

STIC-Biotech/ChemLib

60225

From: Rao, Manjunath N.
Sent: Wednesday, February 13, 2002 7:37 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/749,972

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10C 01
Phone: 306-5681

Date: 2-13-02

Please search the following as soon as possible for application with serial number **09/749,972**

SEQ ID NO: 1 against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO: 2, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
U. S. Patents and Trademark Office
Art Unit 1652, 10A11
Phone: 703-306-5681

Edward Hart
Technical Info Specialist
S. C. / 12/15/02
Conf 12/15/02 Tel: 303-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 2/13/02
Date Completed: 2/15/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 1
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 04
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 22:13:42 ; Search time 1339.35 Seconds
(without alignments)
7870.756 Million cell updates/sec

Title: US-09-749-972-1

Perfect score: 639

Sequence: 1 agtcaaaaggatttttagt.....gaatgggttgccaaatga 639

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pl:**

25: em_ro:**

26: em_sts:**

27: em_sy:**

28: em_un:**

29: em_vi:**

30: em_htgo_hum:**

31: em_htgo_inv:**

32: em_htgo_rod:**

33: em_htg_hum:**

34: em_htg_inv:**

35: em_htg_rod:**

36: em_htg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	635.8	99.5	12815	1	AE007398 Streptococcus
2	635.8	99.5	111135	2	SENEU1906 Streptococcus
3	233.8	39.7	10183	1	AE005502 Streptococcus
4	246.6	38.6	12560	1	AE006276 Streptococcus
5	227.4	35.6	3466	1	AF178425 Streptococcus
6	187.2	29.3	300150	1	AP001507 Streptococcus
7	166.6	26.1	180136	1	BAC180K Streptococcus
8	166.6	26.1	213080	1	BSUB0001 Streptococcus
9	131	20.5	301050	1	AP003130 Streptococcus
10	131	20.5	343590	1	AP003359 Streptococcus
11	109.6	17.2	971	1	MC109 Streptococcus
12	107.2	16.8	49617	6	AX067453 Streptococcus
13	103.2	16.2	10471	1	AE004722 Streptococcus
14	89.6	14.0	12333	1	AE001292 Streptococcus
15	86.2	13.5	16207	1	AE002314 Streptococcus
16	83.2	13.0	311321	1	NMA322491 Streptococcus
17	82.8	13.0	347660	1	AP002994 Streptococcus
18	81.6	12.8	10075	1	AE002421 Streptococcus
19	81.6	12.8	349980	6	AX044030 Streptococcus
20	77.6	12.1	11763	1	AE008073 Streptococcus
21	77.6	12.1	13037	1	AE001769 Streptococcus
22	76.4	12.0	40544	1	SCH5 Streptococcus
23	73.4	11.5	13594	1	AE006204 Streptococcus
24	72.6	11.4	14530	1	AE004276 Streptococcus
25	71.4	11.2	327650	1	MPULM01 Streptococcus
26	71	11.1	14567	1	AE000716 Streptococcus
27	70.8	11.1	9932	1	AE001612 Streptococcus
28	70.8	11.1	11176	1	AE002210 Streptococcus
29	68.2	10.7	633	6	AR139594 Streptococcus
30	68.2	10.7	10238	1	U32728 Streptococcus
31	66.4	10.4	329100	1	SME531787 Streptococcus
32	64.6	10.1	11010	1	AE001873 Streptococcus
33	62.2	9.7	2658	1	LLARCAB Streptococcus
34	56.6	8.9	12059	1	U39679 Streptococcus
35	55.6	8.7	11308	1	AE002101 Streptococcus
36	55.4	8.7	495	1	YEN270424 Streptococcus
37	54.2	8.5	495	1	YEN270427 Streptococcus
38	52.8	8.3	279110	1	RFX03 Streptococcus
39	52.6	8.2	495	1	YEN270429 Streptococcus
40	52.4	8.2	1694	1	AF065312 Streptococcus
41	52.2	8.2	495	1	YEN270423 Streptococcus
42	52.2	8.2	495	1	YEN270425 Streptococcus
43	52.2	8.2	6109	1	AF294440 Streptococcus
44	50.6	7.9	495	1	YEN270426 Streptococcus
45	49	7.7	495	1	YEN270428 Streptococcus

ALIGNMENTS

RESULT 1

AE007398

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AE007398 12815 bp DNA BCT 20-JUL-2001
Streptococcus pneumoniae section 81 of 194 of the complete genome.
AE007398 AE005672
AE007398.1 GI:14972397

Streptococcus pneumoniae.

Streptococcus pneumoniae

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

1 (bases 1 to 12815)

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,

Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J.,

Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,

Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,

Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L.,

McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,

Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,

Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and

TITLE	Fraser,C.M.
JOURNAL	Complete genome sequence of a virulent isolate of Streptococcus
MEDLINE	pneumoniae
PUBMED	Science. 293 (5529). 498-506 (2001)
REFERENCE	21357209
AUTHORS	11463916
	2 (bases 1 to 12815)
	Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
	Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
	Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,R.J.,
	Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
	Khouri,H.A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
	Feidlyum,T.V., Angluoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
	Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
	Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
TITLE	Direct Submission
JOURNAL	Submitted (29 JUN-2001) The Institute for Genomic Research, 9712
FEATURES	Medical Center Dr, Rockville, MD 20850, USA
source	Location/Qualifiers
	1. .12815
	/organism="Streptococcus pneumoniae"
	/db_xref="taxon:1313"
gene	complement(121. .930)
CDS	/gene="SP0923"
	complement(121. .930)
	/gene="SP0923"
	/codon_start=1
	/transl_table=11
	/product="Cof family protein"
	/protein_id="AAK75047.1"
	/db_xref="GI:14972398"
	/translation="MIKLLALDMGDTLLNEAKEIPQAHTAIHKAIEGVKVLQCTGR
	PLFGVLPYKKLELDQNEYIIVNGCSTHOTSQSLVDWKLSPADIEVLVDLAEKS
	DVQLTLDSEHYFLGDKPNQVIENTDAKLVESDLTSLSEEAATSGKFRQCFMELGTK
	BOTDDFQRAEELCORFSGVRSQPIYEAMPGLGTTKATALSRLAEILKIDSEIMAM
	GDANDNIELQFAGLGIAFGNASDIYKSLADAVTSSNEEDGVARAIEKYIL"
	1202. .1423
gene	/gene="SP0924"
CDS	1202. .1423
	/gene="SP0924"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAK75048.1"
	/db_xref="GI:14972399"
	/translation="MILMTKNINLTNEELELIOGGADPGKKNPNGRDWEIEFPVLTLL
	VHGFCPRGTYDSGVIGGNHLCGSAARF"
	1477. .2148
gene	/gene="SP0925"
CDS	1477. .2148
	/gene="SP0925"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AAK75049.1"
	/db_xref="GI:14972400"
	/translation="MKKRAIQILLALSILFYKSTFWRLFNVLAKPYLPASREFEQIL
	LMQSGVLFNAVILVIFAGKRIHFHKQWIRFYILLGYIISYMSDFLFSYLSLS
	NQSLNTEVPMRGQEPFPLVITCFIAPAEELIYRGVLMTFFPKNSPWGDVLLISA
	LIIFYIHINFALPLAFFIYASGLLILALLYRTMKNLYYPILVHILINITAFWDVWLL
	LFSGS"
	complement(2200. .2397)
gene	/gene="SP0926"
CDS	complement(2200. .2397)
	/gene="SP0926"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAK75050.1"
	/db_xref="GI:14972401"
	/translation="MFIFYPQLLYLFIPIYFRSPKPKKKEHPHLLNLYLHPFSPQSP
	FQNFQFKKITGYTPLQYRKEKR"
	2390. .3298
gene	/gene="SP0927"
CDS	2390. .3298
	/gene="SP0927"
	/codon_start=1
	/transl_table=11
	/product="transcriptional regulator, LysR family"
	/protein_id="AAK75051.1"
	/db_xref="GI:14972402"
	/translation="MNIQQLRYVVAIANGSTTFREAAEKMYVSPQSISISVRDLEKELG
	FKPIFRSSGTFTRRQMEFYKSEQLVGFDFIQNYANPEEKDEFVSVAHQHDFL
	PPPTAFESRYPDYKNPRIFESTVQILDEVAQGHSEIGIILNNOKKGIMORVKL
	GLEVILPEFTHYILREGHPLAQKELVEMDLADLPTRFTQEKDEVILYSENFVDT
	SASSOMENVTRDATINGILERTDAYATGSGFLDSDSVNGITVIRLKDNLDRMYVYKR
	EEVELSQAQHLFVEVMQEYFDQRRKS"
	3295. .3756
gene	/gene="SP0928"
CDS	3295. .3756
	/note="identified by match to PFAM protein family HMM
	PF01252"
	/codon_start=1
	/transl_table=11
	/product="lipoprotein signal peptidase"
	/protein_id="AAK75052.1"
	/db_xref="GI:14972403"
	/translation="MKKRAIVAVIVLLILGLDLQVKSIVQOIPLGEVRSWIPNFVSL
	TYLONRAAFSIILOQDLQDLFAVITLVVVGAIWYHLKHMEDSFMMVGLUTLIAGGLG
	NFDIRVSGQGVFDMFHLDFTNFNFAFNVADSYLTGVGVIIILIAMLKEINGN"
	3746. .4633
gene	/gene="SP0929"
CDS	3746. .4633
	/gene="SP0929"
	/codon_start=1
	/transl_table=11
	/product="ribosomal large subunit pseudouridine synthase
	D"
	/protein_id="AAK75053.1"
	/db_xref="GI:14972404"
	/translation="MEIKIETGLRLDKALSLSLSRSLANEQIKSGOVLVNGQVKK
	AKTYQEGDVYTHVPEPEVLEYVAEDLPLEIVYQDEDAVAVNKGPMGVVHPHSAGTTS
	GTVLNMYHIKDLSGINGVRPGIVHRIKDQDTSGLLMIKANDDAHLAQLERGLKOKKS
	LRKYVAIVHGNLNDRGVIEAPIGRSEKDKKQAVTAKGKPAVTRFHVLERGDDYSLV
	ELOLEGTGRHQIRVHMAYIGHPVAGDEVYGPRTKLGHGQFLHAKTLGTFTHPTGKTL
	EFRADIPETFKETLERURK"
	4636. .6519
gene	/gene="SP0930"
CDS	4636. .6519
	/gene="SP0930"
	/codon_start=1
	/transl_table=11
	/product="choline binding protein E"
	/protein_id="AAK75054.1"
	/db_xref="GI:14972405"
	/translation="MKKKLTSLTSLVCAFLGLSWYGNVQAQESSGNKKIHFINVQEGSD
	AIILENSGFPAWDTGEDDFPDCSDSRYPVREGIETSKYKHVITRDRFRLKELGVQK
	LDFILVTHSHDHIGNDELLTYPVDRVYLRKYSDSRITNSERLNDNLYGYDKVLQQT
	AKKGYSVIQNIQTQGAHFQFGDMQILQYENETDSSELKKIWDNSNSLIISVYK
	NKKYIKGLDNDVHGAEDKYGPLIGKVDLMKFNHHHTDKNTKDFIKNLSPSLIVQ
	TSDSLPMKNGVDSEYVNLKRGIERINAASKDYDATVDIRKDGCVNISTSYKPIPS
	FOAGWHKSAQWNMYQAPDSTGEYVAGNIEGEWYFYFNOTGILLONQWKKNNHWFY
	LTDSGASAKNNKKIAGIWIYFNKRNQMEIGWQDKQEWYILDVDSMK7GWLQTMGQW
	YIFAPSGMKMGWKDKETWYNDSTGVMTKEIEVAGOHYYLEDSGAMWGHKKKAK
	DWYFYKTDGSRVAGDKDKWYFLKNGQLLVNGKTPGTYTVDGSGAWLVQVSTEKS
	ATIKTTSHSSEIKESKEVVKDLNENKTSQHSVTFNSTSQDLTSSTSQSSETSVNKS
	SEQ"
	6620. .7729
gene	/gene="SP0931"
CDS	6620. .7729
	/gene="SP0931"
	/codon_start=1

```
/transl_table=11
/product="glutamate 5-kinase"
/protein_id="AAK75055.1"
/db_xref="GI:14972406"
/translation="WKYKRIVKFGTSSLTNEDGSLRSKVKDITQOLAMLEHAGHEL
ILVSSGAIAGCAIGCFKKRPTKIADKQASAAVGGGLLEEVYTNLLRLQIVSAQILL
TODDFVDRKRYRNAQALSVLLNRGAIPINENDSVVIDELKVGNDLUSQAAMVQ
ADLLFLTDVGLYTCNPNSDPRAKRLERETINREIIDMAGAGSSNGTGMGLTKIK
AATIAATESGPYIICSLKSDSMIEAAEETEDGSYFAQEKGLRTQKWLAFYAOSQ
SIWVKGAALSOYKSLLSGIVAEAGVFGYGDIVTFVDESKGSLKGRVQFGAS
ALEDLRSOKAKGVLIYRDDWISITPEIQLLTFE"
7739..9001
/genes="SP0932"
7739..9001
/genes="SP0932"
/codon_start=1
/transl_table=11
/product="gamma-glutamyl phosphate reductase"
/protein_id="AAK75056.1"
/db_xref="GI:14972407"
/translation="MYSRQEQFQVQAVKKSINTASEEVKNQALLAMADHLVAATEEI
LAANLDMAAKGISDVMLDRLYLADRIEAMRGIREVWALPDPIGVLETSOLEN
GLVITKRVAMGVIGIIVESRPNVTSDAALTLKSNVAVLSRGDAYQTTIAIVTAL
KGLLETTIHPNVIQVEDTSRESSYAMKAKGYLDLLIPRGAGLINAVVENAIVPV
IETGTIVHVYVDKDAEDKALSIINNAKTSRPSVCNAMEVILLVHKNKAASFLPRLEQ
VLVAERKEAGLEPIQIRLDSKASQFVSGQAETQDFTEFLDYVLAVKYVSSLEAVA
HIESHTHSIDAIVTENAEAAAYFTDQVDSAAVYVNASTRFTDGGQFGLGCEWGISQ
KLHARGMGLKELTSKYVYVAGDQIRE"
9005..9802
/genes="SP0933"
9005..9802
/genes="SP0933"
/notes="identified by match to PFAM protein family HMM
PF01089"
/codon_start=1
/transl_table=11
```

```
Query Match          99.5%; Score 635.8; DB 1; Length 12815;
Best Local Similarity 99.7%; Pred. No. 1.8e-168;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcaaaaggatttttagtctctctgtgagggaccgagggagcagcagaccagtgtt 60
Db 9918 ATGTCAAAAGGATTTTGTCTCTCTGTGAGGGACCGAGAGGAGCAGCAAGCAGTGT 9977

QY 61 tttagagctctgtaccatttttagaggaagaaaggattagagtggttgacagaccgtgaa 120
Db 9978 TTAGAGGCTCTGTACCAATTTTAGAGGAAAAAGGAGTAGAGGTGTGAGCAGCCGTGAA 10037

QY 121 cctggcggagctgtgattggggagaagattcgggaagtgatttttgatcccaagtcatact 180
Db 10038 CCTGGCGGAGTCTTGATTGGGGAGAAGATTCGGGAAGTGATTTTGGATCCAAATCATACT 10097

QY 181 cagatggatgctaaacagagctactctctctatatgtccagtcgcagacagcatttggtg 240
Db 10098 CAGATGGATGCTAAAAACAGAGCTACTTCTCTATATTGCCAGTCGCAGACAGCATTTGGTG 10157

QY 241 qaataaattctccagcccttgaaagctggcaagtttgggtcatcatgatgattgtttatcgat 300
Db 10158 GAAAAAGTCTTCCAGCCCTTGAAGCTGGCAAGTTGGTCATCATGTGATCGTTTATTCGAT 10217

QY 301 agttctgttgccctatcaggagatttggctgtggtttagattatgaagccattgactggctc 360
Db 10218 AGTTCYGTGGCTATCAGGGATTTGGTCGTGGCTTGATATTAAGCCATTGACTGGCTC 10277

QY 361 aatcagtttgcgacagatgggctcaaacccgatttgacactctattttgacatcagggtg 420
Db 10278 AATCAGTTTGCAGACAGATGGCCFCAAAACCGATTTGCACACTCTATTTTGACATCGAGGTG 10337

QY 421 gaagaaggctggctgattgctgcgaatagtgacccgagagtttaactcgtttggatttg 480
Db 10338 GAAGAAGGGCTGGCTGATTGCTGCTAATAGTAGTGACCGGAGGTTAATCGTTGGATTG 10397
```

```
QY 481 gaagggttgactgcataaaaaagttcgtcaaggtcacctttctctcttgataaagag 540
Db 10398 GAAGGGTTGACTTGCATAAAAAAGTTCGTCAAGGCTACCTTTCTCTCTTGTGATAAAGAG 10457

QY 541 ggaatcgcattgtcaagattgatgtctctcccttggagcaagtttggaactacc 600
Db 10458 GGAATCGCATTTGCAAGATTGATGTCTCCCTTTGGAGCAAGTTGTGGAACTACC 10517

QY 601 aaggctgtctgtttgaacgaatgggcttggccaaatga 639
Db 10518 AAGCTGTCTTGTTCACGGAATGGCTTGGCCAAATGA 10556

RESULT 2
SPNEU1906 SPNEU1906 111135 bp DNA HTG 11-JUL-2001
LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
DEFINITION in ordered pieces.
ACCESSION AL449928
VERSION AL449928.1 GI:11545153
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae.
REFERENCE 1 (bases 1 to 111135)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francisco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae
JOURNAL type 19F clinical isolate
MEDLINE Microb. Drug Resist. 7 (2), 99-125 (2001)
REFERENCE 2135329
AUTHORS 2 (bases 1 to 111135)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francisco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
Direct Submission
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..111135
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 32875 a 20454 c 25072 g 32680 t 54 others
ORIGIN

Query Match          99.5%; Score 635.8; DB 2; Length 111135;
Best Local Similarity 99.7%; Pred. No. 2.5e-168;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcaaaaggatttttagtctctctgtgagggaccgagggagcagcagaccagtgtt 60
Db 73721 ATGTCAAAAGGATTTTGTCTCTCTGTGAGGACCGAGGAGCAGCAAGCAGTGT 73780

QY 61 tttagagctctgtaccatttttagaggaagaaaggattagagtggttgacagcccgtaa 120
Db 73781 TTAGAGGCTCTGTACCAATTTTAGAGGAAAAAGGAGTAGAGGTGTGACAGCCCGTAA 73840

QY 121 cctggcggagctgtgattggggagaagattcgggaagtgatttttgatcccaagtcatact 180
Db 73841 CCTGGCGGAGTCTTGATTGGGGAGAAGATTCCGGGAAGTGATTTTGGATCCAAATCATACT 73900

QY 181 cagatggatgctaaacagagctactctctatatgtccagtcgcagacagcatttggtg 240
```

||||| 73901 CAGATGGATGCTAAACAGAGCTACTTCTATATATGCGATCGACACATTTGGTG 73960
Qy 241 gaaaaagtctctcagccctgaagctggcaagtgtgcatcatcgatcgcttttcatcgat 300
||||| 73961 GAAAAGTCTTCAGCCCTTGAGCTGGCAAGTTGGTCATCATGGATCGTTTATCGAT 74020
Qy 301 agttctgtgctatcagggtattgtgctggtgcttagatattgaagccattgactggctc 360
||||| 74021 AGTTCTGTGCTATCAGGAGTTTGGCTGGCTTAGATATTGAAGCATTAAGCTGGCTC 74080
Qy 361 aatcagittggacagatgggtctaaacccgatttgacactctatttgaacatcgaggtg 420
||||| 74081 AATCAGTTTGGCAGAGATGGCTCAAAACCCGATTGACACTTATTTGACATCGAGGTG 74140
Qy 421 gaagaaggctggtctgtattgctgcgaatagtgacgcgaggttaatcgtttgatttg 480
||||| 74141 GAAGAAGGCTGGCTGTATGCTGCTAATAGTACCGCAGAGTTAATCGTTTGGATTG 74200
Qy 481 gaagggttgactgcatataaaaggctcgtcaaggctaccttctctctggtataaagag 540
||||| 74201 GAAGGTTGGACTTGCATAAAAAAGTTCGTCAAGGCTACCTTCTTCTCGATAAAGAG 74260
Qy 541 ggaatcgattgcaagattgatgtagtctcccttggagcaagttgtggaactacc 600
||||| 74261 GGAATCGCATTTGCAAGATTGATGCTAGTCTCCCTTTGGCAAGTTGTGGAACATACC 74320
Qy 601 aaggctgcttctgttcgcaagtggtctggtcccaaatga 639
||||| 74321 AAGGCTGCTTGTGTTGACGAATGGCTTGGCCAAATGA 74359

RESULT 3
AE006502
LOCUS AE006502 10183 bp DNA BCT 01-JUN-2001
DEFINITION Streptococcus pyogenes M1 GAS strain SP370, section 31 of 167 of
the complete genome.
ACCESSION AE006502 AE004092
VERSION AE006502.1 GI:13621639
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Streptococcus pyogenes M1 GAS.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10183)
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE 11296296
PUBMED
REFERENCE 2 (bases 1 to 10183)
AUTHORS Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA
FEATURES
source 1..10183
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
115..744
/gene="upp"
/note="SPy0392"
115..744
/gene="upp"

/EC_number="2.4.2.9"
/note="Best Blastp hit = sp|P36399|UPP_STRSL URACIL
PHOSPHORIBOSYLTRANSFERASE (UMP PYROPHOSPHORYLASE)
(UPRTASE) >gi|347998|gb|AAA26890.1| (L07793) uracil
phosphoribosyltransferase [Streptococcus salivarius]"
/codon_start=1
/transl_table=11
/product="putative uracil phosphoribosyltransferase"
/protein_id="AAK33430.1"
/db_xref="GI:13621640"
/translation="MGKCVISHPLIQHKLILRROTTSTKDFRELNVETAMLMGVEV
SRDPLEVDVDTQTPVSKYQKOLAGKLAIVPILRAGICGMVDGLLSLVPAKVGCHGM
YRNEETLEPEVYLKLPEDINQRIELVDPMLATGSAILLAVDSLKKRGRANIKFVCL
VAAPGVKKLQEAHPDIDIFTAALDHLNHEHGVIYVPLGDAGDLRFGTK"
713..847
/gene="SPy0393"
713..847
/gene="SPy0393"
/note="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33431.1"
/db_xref="GI:13621641"
/translation="MLVIVLQGNKRDNHQIVAFLGQSFLILFKKSERFLFFKGKASL
"
968..1558
/gene="clpp"
/note="SPy0395"
968..1558
/gene="clpp"
/EC_number="3.4.21.92"
/note="Best Blastp hit = sp|P36398|CLPP_STRSL
ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT
(ENDOPEPTIDASE CLP) >gi|347999|gb|AAA26891.1| (L07793)
ATP-dependent protease proteolytic subunit [Streptococcus
salivarius]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent protease proteolytic
subunit"
/protein_id="AAK33432.1"
/db_xref="GI:13621642"
/translation="MIPVVIETQTSRGSYDIYSRLIKDRILMLTGPVEDNMANSVIA
OLLFLDAQDNTKIDLYVNTPGGSVAGLAIVTMFIRADVOTIVMGMAASMGTVIA
SSGTGKGRMLPNAEYIHPQMGSGTGTQOTDMAIAAEHLLKTRHLEKILQAQNAK
TKQIHKDAERDYWMSAEETLAYGFIDEIMENNELK"
2050..2325
/gene="SPy0397"
2050..2325
/gene="SPy0397"
/note="Best Blastp hit = pir|D69874 conserved
hypothetical protein ylbG - Bacillus subtilis
>gi|2340003|emb|CAB11353.1| (298682) ylbG protein
[Bacillus subtilis] >gi|2633871|emb|CAB13373.1| (299111)
similar to hypothetical proteins [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33433.1"
/db_xref="GI:13621643"
/translation="MFQKQERIGLVVLYYNRDARKLSKFGDLYHSHKRSRYLIYIN
KNDLTKLEMRRLKCVKDIRPSAFDDIDRQFVGNLHRDETNNHQG"
2574..3209
/gene="tmk"
/note="SPy0399"
2574..3209
/gene="tmk"
/note="Best Blastp hit = dbj|BA803761.1| (AP001507)
thymidylate kinase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative thymidylate kinase"

```
/protein_id="AAK33434.1"
/db_xref="GI:13621644"
/translation="MITGKLITVEGPDGAGKTTVLQELPLLKQKVAQDILLTTREPGG
VAISEHIRELIDINHTAMPDKTELLYTAAROHLEVKVLPALAEAGOLVIFDRFIDS
SVAYOCAGRLIKADIQWLNEPATDGLPDLTLYFDVPEIGLARINAOORENRLD
LEFIEIHQVRKGYLALAKEHPKRIVTIDATKPLKEVSVVALEHVLALLA"
3227..4102
/gene="SPy0400"
CDS
3227..4102
/gene="SPy0400"
/notes="Best Blastp hit = dbj|BAB03763.1| (AP001507) DNA
polymerase III delta' subunit [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative DNA polymerase III delta' subunit"
/protein_id="AAK33435.1"
/db_xref="GI:13621645"
/translation="MDLAOKAPNVYQAFQITLKKRLNHAYLFESGFANEMALFLAK
VFCEKQKQTPGCHRSQOLIEQGFADVTVEPTQGVIKTDVVREMANFVSQTGE
NRQVFIIDKDKMHINAAANSLIKYIEEPOGEAYIEFLNDDNKVLPTTKSRQVQFQ
PKNEAYLYLQAOEKLINHOAKLIVAKLATNLSHLRLQTSKLELITQAREVSIWL
KDQQLAYLALNRLVQLATEKEQDLVLTLLTLARERAQTPLTQEAQVQARLMWQS
NVNFTLEVMVNSE"
4121..4360
/gene="SPy0401"
CDS
4121..4360
/gene="SPy0401"
/notes="Best Blastp hit = dbj|BAB03764.1| (AP001507) signal
peptidase-like protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33436.1"
/db_xref="GI:13621646"
/translation="MGRKSLSKDLPVATKVAHKKKAPENTAFATDLMKEYSRLIT
QNQLEMTFVNSVFLLDQHVVLIPFSAEHRMDFNC"
4765..5088
/gene="SPy0405"
CDS
4765..5088
/gene="SPy0405"
/notes="Best Blastp hit = dbj|BAB03765.1| (AP001507)
unknown conserved protein in B. subtilis [Bacillus
halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33437.1"
/db_xref="GI:13621647"
/translation="MNKELFADFQFSQNLMTVLAEIEMKKOVOSLSVEENTILRLE
NTKLRLRSLHLEHTVAKNPSKQRKDHLGIYDEGFHICNFFYQORRENDEECMFQRE
LLDRK"
5093..5956
/gene="SPy0406"
CDS
5093..5956
/gene="SPy0406"
/notes="Best Blastp hit = dbj|BAB03768.1| (AP001507)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33438.1"
/db_xref="GI:13621648"
/translation="MOVOKSFDKKTSGLTLYLPTPIGNLQDMTERAVATLKVEDFIC
AEDTRVTGLLKHFDIATKOISFEHNAYEKIPOLDLLISGRSLAQSDAGNPSISD
PGHDLVKAISDIAVVALPGASAGITALIAGLAPQPHVYFGLPKRAGQQAFFED
KHHPYQMTFYESPYRKDTLTNNLACYGDQVVLVREFTKLFEEYQSGISILSYL
EETPLKGECLLIIVAGAQAQSEVELTADVDLVSLVQKEIQAGAPNOAIKTIKAYQVN
RQELYQQFHD"
5983..6375
/gene="SPy0407"
CDS
5983..6375
/gene="SPy0407"
/notes="Best Blastp hit = emb|CAA61744.1| (X89566)
```

```
cytochrome-c oxidase sbunit I [Nitrobacter winogradskyi]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33439.1"
/db_xref="GI:13621649"
/translation="MTEDQATILILGVYIILAVFLLSALLCLIANRRLPKRAGYKSW
YSLIPFYAGFVKHITFGEDKNWFIYFIWYIAIYVYVHFCYVRAFGSKGLAVFGLF
FPGIASLIVAFGKTYTTECHPVPHLLN"
complement(6422..7051)
/gene="SPy0408"
CDS
complement(6422..7051)
/gene="SPy0408"
Query Match 39.7%; Score 253.8; DB 1; Length 10183;
Best Local Similarity 65.4%; Pred. No. 1.le-60;
Matches 388; Conservative 0; Mismatches 202; Indels 3; Gaps 1;
Qy 1 atgtcaaaagagatttttagtctctcttgaggaccagaggagcagcagcaagaccagtgtt 60
Db 2574 ATGATCACAGGAAAACATCAATACCGTTGAAGTCCAGATCGCGACGGAACGACTGTT 2633
Qy 61 ttgagggtctgtctaccaatttttagagggaaaagagtagagg---tgttgaccacctgt 117
Db 2634 TTAGAGCAACTCATTCACACTGTTAAACAAAAAGTGGCTCAGGATATCTTGACAAACA 2693
Qy 118 gaacctggcaggtcttgattgggggagaagattcgggaagtgtatttgatcccaagtc 177
Db 2694 GAGCCTGGAGGTGTTGCCATTTCAGAACACATAGAGAACTATTTTAGACATCAACCAT 2753
Qy 178 actcagatggatgctaaacacagctactctctatatatgcccagtcgcagcagcagcttg 237
Db 2754 ACTGCTATGGACCCCTAAACAGAACTCTTACTTTATATGCGGCTCGAAGGCAGCATTTG 2813
Qy 238 gtggaaaaagttcttccagcccttgaagctggcaagttgggttcacatcgatcgcttttct 297
Db 2814 GTGGAAAAGGTACTACCAAGCTCTTGAAGCAGCCAGTTGGTTTATTGATCGTTTCATC 2873
Qy 298 gatagttctgtgctatcagggtatttggtcgtggtcttagatattgaagcactgactgg 357
Db 2874 GATAGTTCAGTAGCTTATCAAGGTGCGGGACGCGGTTTGTGCAAGAGACATTCAGTGG 2933
Qy 358 ctcaatcagttgcgacagtggtctcaaacccgatttgacactctatttgacatcgag 417
Db 2934 CTCACAGGTTCCGAACAGATGTTTAGAGCCTGACTTGACCCCTTATTTGATGTTCTT 2993
Qy 418 gtgaaagaagggtcggtcgctattgctgcgaatagtgaccgcgaggttaactcggttgat 477
Db 2994 TCTCAAAATTGTTTAGCAGCGATTAAACGCGAACCAACAAAGAGAAGTCAATCGTTTAGAT 3053
Qy 478 ttgaaagggttggtgactgcataaaaaagttcgtcaagctacccttctctctgataaaa 537
Db 3054 TTGAAAACAATPAGAGATACATCAACGAGTAGCAAGAAAGGATACCTAGCCCTTAGCTAAGGAG 3113
Qy 538 gagggaaacgattgctcaaatgattgctgctctccctcttgaggcaagtgtt 590
Db 3114 CATCAAAAACGATTGTGCACAAATTGATCGGACTTAACACCTAAGAAAGATTGT 3166
RESULT 4
AE006276
LOCUS
DEFINITION Lactococcus lactis subsp. lactis IL1403 section 38 of 218 of the
complete genome.
ACCESSION AE006276 AE005176
VERSION AE006276.1 GI:12723260
KEYWORDS
SOURCE Lactococcus lactis subsp. lactis.
ORGANISM Lactococcus lactis subsp. lactis
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
```

REFERENCE
AUTHORS

1 (bases 1 to 12560)

Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.

TITLE

The Complete Genome Sequence of the Lactic Acid Bacterium

JOURNAL

Lactococcus lactis ssp. lactis IL1403

MEDLINE

Genome Res. 11 (5), 731-753 (2001)

PUBMED

21235186

REFERENCE

11337471

AUTHORS

2 (bases 1 to 12560)

Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.

TITLE

Direct Submission

JOURNAL

Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de

FEATURES

Vilvert, Jouy en Josas 78352, France

source

Location/Qualifiers

1. .12560

gene

/organism="Lactococcus lactis subsp. lactis"

CDS

/strain="IL1403"

/sub_species="lactis"

/db_xref="taxon:1360"

129. .320

gene

/gene="ydljB"

/note="L195067"

CDS

129. .320

gene

/gene="ydljB"

/note="EVIDENCE BY GENEMARK BIO16.01 UNKNOWN. General.

homologs not detected in Streptococci."

CDS

/codon_start=1

/transl_table=1

/product="UNKNOWN PROTEIN"

/protein_id="AAK04486.1"

/db_xref="GI:12723261"

/translation="MIHELTDRAHQHFSTRKIDAFWIFEPTTGALFEGGLVSENKAI

TFEINAKTLEVFENEKIL"

362. .679

gene

/gene="trxB"

/note="L195348"

CDS

362. .679

gene

/gene="trxB"

/note="EVIDENCE BY HOMOLOG BIO02.10 Thioredoxin,

glutaredoxin, and glutathione. similar to the thioredoxin

H-type from plants."

CDS

/codon_start=1

/transl_table=1

/product="thioredoxin H-type"

/protein_id="AAK04487.1"

/db_xref="GI:12723262"

/translation="MIIPENIENLAELVKGPEKTVFFFTAGWCDCNFIKPKMPEIEV

ENPDFKEVDREIDMLAIEWGIPSEFVIEDGKEKARLVNKLKRTKEEVNTFLA

SAK"

765. .1391

gene

/gene="ydljD"

/note="L195751"

CDS

765. .1391

gene

/gene="ydljD"

/note="EVIDENCE BY HOMOLOG BIO15.01 HYPOTHETICAL. homolog

phenylalanine tRNA ligase"

/codon_start=1

/transl_table=1

/product="HYPOTHETICAL PROTEIN"

/protein_id="AAK04488.1"

/db_xref="GI:12723263"

/translation="MIATYNTNVGDVLMILITANDKGEKVNFRKGVARVFIETGAT

VANNIFEAKSLLANLNGQVWLSSDDIETLNKELGSGFSDVLVHDAQKFFVVAEIV

EMEDHPSDHLHCKVNVGFADPQIVCGAPNAALGLTKTVAALPGAMNPGSLIFPGA

LRGVQSYGMCLCSARELPNAPVRGIIELSSDLSAGKAFDSETMWOA"

1554. .2894

gene

/gene="noxE"

/note="L196579"

CDS

1554. .2894

gene

/gene="noxE"

/note="EVIDENCE BY HOMOLOG BIO06.01 ENERGY METABOLISM.

Aerobic"

/codon_start=1

/transl_table=1

/product="NADH oxidase"

/protein_id="AAK04489.1"

/db_xref="GI:12723264"

/translation="MKIVVIGTNGHAGIATANTLIDRYPGHIEIVMIDRNSNMSYLGCGT

AIWVGRIEKPDELFYKAEDFERKGVKILPTEVESEIDFTNKKIYAKSKTGEKITES

YDKLVLATGSRPIIPNLPGDKLGHFLKFOEQADIDEEFAKNDVKRIAVIGAGYIG

TEIAEAAKRRGKVELLPDAESTLSIYVDEEFKAGMDENLAQHGLHFLHGLAQEFKA

NEKHVSQIVTNKSTYDVLVINCIGTANSALAGEHLETFKNAIKVKDKHQQSSDDPD

VSAGDVATIYSNAALQDFTYIATALASNAVRSGIVAGHNHIGGKSIESVGVGSGNISIFG

YNTSTGLSKAAKIKGLEVSFDFEDKQAWFLHNNDSVKIRIVYETKRRRIIGAQ

LASKSEIIAGNIMNFSIAIQEKKTIDELALDLFLPHFNSPYNYMTVAALNAK"

2984. .3373

gene

/gene="ssbA"

/note="L0299"

CDS

2984. .3373

gene

/gene="ssbA"

/note="EVIDENCE BY HOMOLOG BIO10.02 DNA replication,

restriction, modification, recombination, and repair. THIS

PROTEIN IS INVOLVED IN DNA RECOMBINATION AND REPAIR"

CDS

/codon_start=1

/transl_table=1

/product="single-strand binding protein"

/protein_id="AAK04490.1"

/db_xref="GI:12723265"

/translation="MNKMLIGRLTSAPETSKTNDKSYVVTLVAVRPFKNEKGERE

ADFSIIWKSATETLVSAYAKKGLSISIEGIRNTYTDKONQHYVTEILGLSLYDLL

ESRATIALRESAITEETLLDAEELP"

3493. .3777

gene

/gene="groES"

/note="L198515"

CDS

3493. .3777

gene

/gene="groES"

/note="EVIDENCE EXPERIMENTAL PMID:8486277 BIO04.03

Chaperones"

/codon_start=1

/transl_table=1

/product="10 KD chaperonin"

/protein_id="AAK04491.1"

/db_xref="GI:12723286"

/translation="MLKPLENVLVRVKEEEKSMGGIVILTSASQEKPTQAEVAVGE

GKTNHHGTLSPLVKVGDVIFEFKFACTVVKMDGEEFLILKDSLLAIVE"

3865. .5493

gene

/gene="groEL"

/note="L198893"

CDS

3865. .5493

gene

/gene="groEL"

/note="EVIDENCE EXPERIMENTAL PMID:8486277 BIO04.03

Chaperones"

/codon_start=1

/transl_table=1

/product="60 KD chaperonin"

/protein_id="AAK04492.1"

/db_xref="GI:12723267"

/translation="MSKDIKFSDDARTAMRGIDILADTVKTTLTPGKRVNVLEKSYG

SPLITNDGVTIAKEIELEHDFENHAKLVSEVASKTNDIAGDGTITATVLTQAIYREG

LKNVTAGNPGVIRGIELAAETAASIKEMAFVHDKSAIAQVATVSSRKEVGEYI

SDAMERVSGDGVITIEESGMQTELDVVEGQDVGSLYSQTMVSNTEKMWAELENDPYI

LITDKKISIQEILPLEQLIKTNRPILLIIVADDVDRGALPTLVLNKIKGVNVVAVKA

PGFGDRRAQOELDIALITGGTVITEELGLDKATLEALGOAAKATVDKDHITTEVEGA

GSADISDVAITKAOIKETKTSDFDREKLOERLAKLAGVAVKVGATTELEKAMKL

LIEDALNAAVVEEIVSGGCTALVNAIALDKLSEGDIOGTGINVRRALEPVRQ

LIANAGVEGVIIDKLRSEVGTGNAATGQVNMIEEGIVDPKAVTRSAQLQNASVA

GLLITTAVVANKEPEAPAMPMPDPSMGGM"

complement(5539. .6351)

/gene="yeaA"

/note="L582"

complement(5539. .6351)

/gene="yeaA"

/note="EVIDENCE BY HOMOLOG BIO15.01 HYPOTHETICAL.

homology to yycJ B.subtilis"

[illegible]

Qy	367	tttgcacagatgggtcctcaaacccgatttgacacctattttgacatcgagggtgaagaa	426
Db	9199	TTTTCAACAGATGGTGTGAAGCCCTGATTGACGCCTTTATTTGATGTTGATACGGATGTT	9258
Qy	427	gggctggctgatattgtgsgaatagttagccgcgcagaggttaatacgtttggatttgaagagg	486
Db	9259	GCCCTTGAGCGAATTAACAATAATTCGAGCTCATGAGGTAAATCCTTTAGATTTTGGAAAAGA	9318
Qy	487	ttagacttcataaaaaagtcgtcaaaggtcaccttctctctctcttggtgataaagaggaaaat	546
Db	9319	GCTGAGATGATCCTGAAGTTCGTGAAGTTATTTGGAGATTGTTGTTAAAAAGAACCTTGAG	9378
Qy	547	cgcatttcgaagattgatctagctctccccttggagcaagttgttgaaactaccaaagct	606
Db	9379	CGTTTGTGAAAAATAGATGCAAGTCAACCTTTAGAAAAAGTGGTAGCTGATACACTTTCC	9438
Qy	607	gtott	611
Db	9439	GTCT	9443
RESULT	5		
AF178425/c			
LOCUS	AF178425	3466 bp	DNA BCT 25-JAN-2001
DEFINITION	Lactococcus lactis TcsCorf1 (tcsCorf1984), Rrc (rrc), and Kinc (kinc) genes, complete cds; and TcsCorf2 (tcsCorf2) gene, partial cds.		
ACCESSION	AF178425		
VERSION	AF178425.1	GI:12483715	
KEYWORDS	Lactococcus lactis.		
SOURCE	Lactococcus lactis		
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus		
REFERENCE	1 (bases 1 to 3466)		
AUTHORS	O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G.F., Ehrlich, S.D. and Morel, P.		
TITLE	Six putative two-component regulatory systems isolated from Lactococcus lactis subsp. cremoris MG1363		
JOURNAL	Microbiology 146 (Pt 4), 935-947 (2000)		
MEDLINE	20244638		
PUBMED	10784052		
REFERENCE	2 (bases 1 to 3466)		
AUTHORS	O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G.F., Ehrlich, S.D. and Morel, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-AUG-1999) Department of Microbiology, National University of Ireland, Cork, Western Road, Cork, Ireland		
FEATURES	Location/Qualifiers		
source	1..3466		
	/organism="Lactococcus lactis"		
	/strain="MG1363"		
	/db_xref="taxon:1358"		
gene	complement(171..806)		
CDS	/gene="tcsCorf1984"		
	complement(171..806)		
	/gene="tcsCorf1984"		
	/note="thymidylate kinase-like"		
	/codon_start=1		
	/transl_table=11		
	/product="TcsCorf1"		
	/protein_id="AAG53731.1"		
	/db_xref="GI:12483716"		
	/translation="MNGILSLLEGPDGAKTTVLQELPEIQKMKREVVPTREPGRVAVGEIQRILDPNKTEIDSKTELMLFAARLRHQEKMLPALRAGKVIVDRISSVAAYGVYQGLIDPVEVDWNTFATDLGLPLTFYFDITDVALERIMKNRADEVRLDLE RAMHRRKREGYLEIVAKEPGREFVKIDASQSLEKKVVADLTILEVLKRKFVSSEF"		
gene	984..1685		
	/gene="rrc"		
CDS	984..1685		
	/gene="rrc"		
	/note="putative response regulator"		
	/codon_start=1		

TITLE
Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125

JOURNAL
MEDLINE
99184646
REFERENCE

AUTHORS
Takami, H. and Horikoshi, K.

TITLE
Analysis of the genome of an alkaliphilic Bacillus strain from an
industrial point of view

JOURNAL
MEDLINE
20263314
REFERENCE

AUTHORS
Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sakiyama, T.,
Hirama, C., Fuji, F. and Takami, H.

TITLE
Characterization and comparative study of the rrm operons of
alkaliphilic Bacillus halodurans C-125

JOURNAL
MEDLINE
20426005
REFERENCE

AUTHORS
Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N.,
Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and
Horikoshi, K.

TITLE
Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis

JOURNAL
MEDLINE
20512582
REFERENCE

AUTHORS
Takami, H. and Takaki, Y.

TITLE
Direct Submission

JOURNAL
MEDLINE
20512582
REFERENCE

AUTHORS
Technology Center, Deep-sea Microorganisms Research Group; 2-15
Natsushima, Yokosuka, Kanagawa 237-0061, Japan
(E-mail: takami@jamstec.go.jp,
URL: <http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html>,
Tel: 81-468-67-3895, Fax: 81-468-66-6364)

FEATURES
source
1. 300150
/organism="Bacillus halodurans"
/db_xref="taxon:86665"
/note="alkaliphile"
584. .1933
/gene="dnaA"
584. .1933
/gene="dnaA"
/note="BH0001"
/codon_start=1
/transl_table=11
/product="initiation of chromosome replication"
/protein_id="BAB03720.1"
/db_xref="GI:10172613"
/translation="MENIHDLWELRALKMEKKVSKPSFETLWTKQTKANSTEDSTIIIT
APNEFARDWLEKHDELISEITDIDLTVGLYPKFVIPTSQLDPEPVEQELKPKMKOPP
AONGEPNNMLNDKYTFDTFVIGSGNRFHAASLAVAEAPAKAYNPLFYGVGLGKT
HLMHAIGHVMDHPNPNKVVYLSSEKFTNEEFNAIRDKNVFNKRYNRVVDVLLIDDI
QFLAGEQEDEFHTFNALHDENKQIVISSORPPKEIPTEDLRKSRFEWGLTIDIT
PDPLETRIALIRKKAENLDIPNVEIYANGIDNFIREDGALIRVAYASLIND
MADLAELAKIDIPNAPKPRVITIDIKTVGEYFHVLEDFKAKRTKSVAPPROIA
MYSRLSDASLPKIGSEFGGRDHTTVIHAHEKISKLLSTDQELQDIQISDKLRS"
2103. .3245
/gene="dnaA"
2103. .3245
/gene="dnaA"
/note="BH0002"
/codon_start=1
/transl_table=11
/product="DNA polymerase III (beta subunit)"
/protein_id="BAB03721.1"
/db_xref="GI:10172614"
/translation="MHFVIDRIFVQNVHNSKAVSSRTTIPILTIKIKIVADHEGVTLL
TGSDSISITETPLPEGRQNVKQGSIVLOAKVFAEIVKLPQEOTEIHVQDSF
VTIRSGSNLNDLPDEYPRLPVLEHDHVRLPQKILKDIROTVEAVSTQETRP
VLITGVNFEIDGILLTCTATDSHRLANRKYVPVKNDELQFSNVNIPGKSLNELSKILD
ENEELDIIVTNENQTLFKLKNMLFFSRLEGKYPVTKNMPKEAKTSFAVHTKAFLOT
LERALLSREGNQVINLKTLDGVDVVEVTAITPEIGKVTENVATQGLEGEELRISFNG

gene
3429. .3650
/gene="BH0003"
CDS
3429. .3650
/gene="BH0003"
/note="BH0003
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB03722.1"
/db_xref="GI:10172615"
/translation="MSQQLPITTSYITLQQLLKEVGVIDTGGMAKWYLAHEVIVNGE
SEDRRGKLFSGDSVTIADGTVIDLVDAA"
3662. .4777
/gene="recF"
3662. .4777
/gene="recF"
/note="BH0004"
/codon_start=1
/transl_table=11
/product="DNA repair and genetic recombination"
/protein_id="BAB03723.1"
/db_xref="GI:10172616"
/translation="MHIERLTQFRNYDELVSFEPNVNIIIGENAQGTNVEIAY
FLAKASHRTARDKELTQWEAPFARIEGAFQKQNGPLSLHVLGSGKGVKXVNGLEQR
RLSDYICAVNVVMPGPDNLNVRGSPQIRRRFLDMELGQSPVLLHQLAMYQKILQQR
NHLKOLFCKANSDDPMLDVLTDOLIEVAVETKRFETIQLQWAEIHOAISRGKE
KLVTYEPSVHVSEOLNLSKLREGFYQAEOKERSEORGTTLGGHRDDLVSFVNDK
DVQYTSQGOORTALSLKLAELMEKTEVGDYPILLDDVLSELDYRSHLLHLAIO
HRVQTVTITNVGDIDHDTLQEAITYEVEQGSUKKKG"
4788. .5057
/gene="BH0005"
4788. .5057
/gene="BH0005"
/note="BH0005
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB03724.1"
/db_xref="GI:10172617"
/translation="MFIHGGDHVIRSKDVIAILDRDMEQSSGVTKYELHHOKNGNI
VTISSDLIKSIVVTPDTIYIYSPISSTLKRSHAVSELDRFGEE"
5130. .7043
/gene="gyrB"
5130. .7043
/gene="gyrB"
/EC_number="5.99.1.3"
/note="BH0006"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BAB03725.1"
/db_xref="GI:10172618"
/translation="WTRQQAYDESOIQVLEGLEAVKRKPGMYIGTSSRGLHLVWE
IVDNSDEAMAGCDGIVIEEDNSITVDNGRIPGVGHEKMGPAVEIVMTVLHA
GKFGDEGGYKVGSLGHGVASVYNALSTMLEVHEHREGKVHYQKFRGVPAADLVIG
TTDRCTKTHFKPDGDIETFTTVFYDTHLASRLAFLNKLRIKIKTDMREEKSD
FYEGGSIASFVEHLNTRTLEHTPIHIEGKQGVYVEIAYQVNDGFTSNLYSFANNI
NTHGGTHESGFKTGLTRVINDYARKHNLKPSDNLGTEDREGTLAISVKKIPDPQ
FEGQTKLGNSEARTITDSLTFEHSRFLIEHPQVARKLVKDLKGLMASRAREAKKAR
ELTRRKSALENSLPGLKADSSRDASISRIYVEGDSAGSAGQGRDRHFQAILPLR
GKILNVKARLDKILANNEIRAITALTGTIGDGFIDEIKARYHKIITMDADVGDGAI
RTLILTFEYMRPLIEHGVIYAQPLKYVQGGKIOYAGDAEIQEILSOLPDKAK
PGIORYKGLGEMNPSQLWETMDPGQRTVLQVLTEDAMKADEIFEILMDRVEPRDRF
IQENARYVKNLDI"
7065. .9566
/gene="gyrA"
7065. .9566
/gene="gyrA"
/EC_number="5.99.1.3"
/note="BH0007"
/codon_start=1
/transl_table=11

CDS

```
complement(5467..5831)
/gene="dnaC"
/codon_start=1
/transl_table=11
/product="replicative DNA helicase"
/protein_id="BAA05176.1"
/db_xref="GI:467330"
/translation="MTDLLNDRLPPONIEAEQAVLGAIFLPQSALTALASEVILPDDFY
RMSHQIKYNAMVLGRGPEVDLVTVTSELANTDLLEEVGSGISYLTIDIANSVPTAANI
EYAKIVEKSLIRLRITATTIAQGYTREDEVDLLSEAEKTIWEVAQRKNTSAFQ
NIKDLVOTYDINEOLYNRKGDITGPTGFTELDRAGFORNDLIIVAARP SVGKTA
PALNAQNAVATKTDSEVALFSLFMGAEOLVMEMLCAEGNINAQNLRTGNLTEDRGKL
TWAMGSLNSGIIYIDTGPGRVSEIRAKCRRLKQESGLGMILDLQLQGGSGRKN
RQGEVSEIRKLSARELQVPVIALSQLSGVEGRQKRPMSDIRESGSIEQADAI
VAFLYRDYDKETENKNIIEIIIAKQRNPGVGTVSFLAFVKEKFNVLNRRERFDAGV
PGA"
```

CDS

```
7001..7201
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05177.1"
/db_xref="GI:467331"
/translation="MIKKYAITPNVDADGWFEIVENVAFTALYTSKDAAIERAKQVAK
ENSPSKLVIYDQFKNVEEHSF"
7447..7713
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05178.1"
/db_xref="GI:467332"
/translation="MHKNPPYLFYNGHPPIALFLCDIFLHFTSDSCTMDRREEGSFVR
PQIQAETAQKLAESLNLPLEIGHMPHQHILLAKMAELQKEDKS"
7786..8994
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05179.1"
/db_xref="GI:467333"
/translation="MPHPHNKKIQSWLITGIIIFIAFNLRPAITSVGPVISSIRAEHL
MNGAAGFTLALPSFAVSLPLAKLQGLGNERTLWLGLVLLIGVLTSTGYTAA
LFFGTALIGVTAIGNVLLPSLKHKYPEKGMISLSTYTSNMNIPAAIASGYSVPLAT
OMNGWKQAFLLWGLLALLIWLIPOLHROTANQTMKLOSSSIWASKMANYVIFM
GLQSLFYSSIAWFFELIKSHGIDTATAGMWSLMQFASLPSTFLPVLADRVKQOGR
IVAAIASVYILGCLLAGGSHLLAINMIIGIGGSSISLALLIGLSRNSAQQA
ALSGMSQFGLYLAAGVIFVGYLFDQTHSWTMPVLLVLLAALVMAAGQAGRDRIYI
FOSEKORNSA"
9099..11156
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05180.1"
/db_xref="GI:467334"
/translation="MKTKRFDAAILILLAAAFNLNTHWDDTANOVYLAAVKSWTO
SFHNFYASDFSPGVYVDKPPVVLWIOFIALLIFGVHTWSVIIIOALAGVSVELLY
RAVKPTFGGAARIALKNMALPIAVAVSRNNIDSMVLFTLLLGSTCLLRVAKGKL
WMLLTAFILGRIAKNMNMMAFWLPAPFVFLIARSVLKKKIGSLLSLLTGLS
LSWA TAVDSASSSRPYGSSQTSNVLEAFNGYNTERLFGOTTGLAKGDMNAAGGN
MONQDMAPNGSSFSFONGNOSHQAOPPPNGOTGALNGGGTPTTGNGPGN
GGPGGGGKSVNMGFTGDPPLRLFQSALSQOISMLPESLIGLLGAILISWYDRGRH
AKKMETLEWALVPVAGFFSIAGFFHOYIILMLAPPIALSGIGWTMHLRYKNKK
DNSSYLLPAVLIIVAFQVYIISAGTTSQIGSVMMYVGLGLGIGITLALLMKRSHPS
KLTTIISLCVLLTPVYSATPLLYXGNSVLPESGPKLKGSGANGNMFSSEYDSGLLS
YLRKNTGSEYLFATLPVTAAPYIIYENESVMAMGGFNGTDPILTVSLKKLIEGK
VKYFLLSENNSGNSELVSWIKKNGKEITSDEYSGSSSTNSVQGMRRGRPGGESQOTLY
LVE"
```

CDS

```
complement(11193..11642)
/gene="rplI"
complement(11193..11642)
/gene="rplI"
/codon_start=1
/transl_table=11
/product="ribosomal protein L9"
/protein_id="BAA05181.1"
```

gene

CDS

```
complement(11193..11642)
/gene="rplI"
complement(11193..11642)
/gene="rplI"
/codon_start=1
/transl_table=11
/product="ribosomal protein L9"
/protein_id="BAA05181.1"
```

CDS

```
/db_xref="GI:467335"
/translation="MKVIEFLQDVKGKKGKVENKVNADGYAHNFLIKKKGLAVEANASNI
SALNGQKQKEKEATAELQAKSLKETLKLKTVLSAKSGEGRGFSVTSKQITQL
OKDHNKVKDKLELPDGRALGYTNVPVKLHPVQAVLKLVHVKEEA"
complement(11639..13618)
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05182.1"
/db_xref="GI:467336"
/translation="MPSFYEKPLFRYPYIALTALSITILISFYFNWILGTVEVLLLA
VILFFIKRADSLIRQIDAYISTLSYRLKKVSEELMEMPIGLMFLNDQYYIEWANPF
LSCSFESTLVGRSLDTCESVPLIKQEVSETVTLNDRKFRVTKRDERLLDFDVF
TQIOIEKLEYERNETVLIAYFLDNYDDVTQGLDDQTRSMNSQVTSLLNAQOEYGI
LKRTSSERTIAVNLHAILTELENSKFSILDEYREKTSFDGVALTSLVSGVGSYSLKE
LGDALSGDULALGRGDOVAIKLPNGKVFYKGNPNMEKTRVRARVISHALKETV
TESSNVIIMGHKFPDMDSIGAAIGILKVAQNNKDGFIVIDPNQIGSSVQRLIGEIKK
YEELWSRFTIPEAEWISNDNLLVIVDTHKPSLWEEERLVNKHIEHVIVDHRGEE
FIRDPULVMEPYASTAEVLTELELQPKRLKINMEATALLAGIIVDTKPSLRKG
SRTFDAASYLRAGADTVLVQKFLKETVDSYIKRAKLIQHTVLKDNIAIASLPNEE
EYFDQVLIQAADSLSMESEASFAVRDEQTVCSARSLSGEVNVQIIMEALEGGG
HLTNAATQLSGISVSEALERLKHAIIDEYFEGGVQR"
complement(13655..14584)
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="BAA05183.1"
/db_xref="GI:467337"
/translation="MKQTRALMEGAILISLFAITLLVVVVPVIGTILLFALPLPML
YTRHGLKUGIWMGAVSLPVFIVSGNGLIVAFMSACAGIAMGHFKRKEPGHALIS
GALLIYSLIVFYVISIQFLGINIDEAMTQYRSLDITVETAKOSGNAGQFQKQLK
MEEQLGIVLYLFTPTAIVMGVIFSFSLYLIATKPLLRFPDIPNLKPFRELKPPQSVV
```

CDS

```
Query Match 26.1%; Score 166.6; DB 1; Length 180136;
Best Local Similarity 55.4%; Pred. No. 6.3e-36;
Matches 322; Conservative 0; Mismatches 259; Indels 0; Gaps 0;
QY 10 ggatttttagtctctcttggagcagagggagcagcagaccagtgttttagagcgt 69
D 102765 GGTTTATTTTACATTCGAAGTCTCTTATATGCGGCTGCAAGACGCCAGCATTTAGTTGAAAAAGTA 103004
QY 70 ctctaccaatttttagaggaagagtagaggtgttgacgacccctgaacctggcgga 129
D 102825 ATCAAAACATCTATGACAGAGAAGGCCCTTCAGGTATATGCAACTCAGACGCCAGCGGA 102884
QY 130 gtcttgattgggagagattcgggaagtatttgatcctcaagtactcagatgat 189
D 102885 ATAGATATTCGACAAACAATATAGAGAGTTATTCTTGAATGAGACAAATATATTAATGAC 102944
QY 190 gctaaacagagctactctctctattgcccagtcgcagacagcatttggtggaaaaagtt 249
D 102945 CCAAAAACAGAGCTCTCTTATATGCGGCTGCAAGACGCCAGCATTTAGTTGAAAAAGTA 103004
QY 250 cttccagcccttgaagctggcaagttgggtcatcatgatgattgtttatcgatagttctggt 309
D 103005 AAACCTGCTTTAGAACAGGGGTTTATCGTTCTTTTTCGACAGATTCATGACAGACGCTCTT 103064
QY 310 gctatcagagatttgctgctgcttagatattgaacccattgactggctcaatcagttt 369
D 103065 CGGTATCAAGGATATGCAAGAGGCTTGGAAATGATGAGGTTCTGTCCATTAATGAATTT 103124
QY 370 gcgacagatgggtcaaacccgatttgacactctattttgacatcgaggtgggaagaggg 429
D 103125 GCCATTGGACATATGATGCCCTTAAACGGTTTATTTTCAATTGATCGGAAGAGGA 103184
QY 430 ctggctcgattgctgcgaatagtgaccgcgaggttaatcgtttggatttgggaaggggttg 489
D 103185 CTGAAGCGTATTACGCAAAACGGCTCACGGGAGAAAAACACAGGCTCGATTTAGAAAAAATG 103244
QY 490 gacttgcataaaaaaagttcgaagctacctcttctcttcttgataaagaggggaaatcgc 549
D 103245 GACTTTTCATAAAAAGTACAAAGAGGTTATCAAGAGCTGTGAAGCGATTTCCCGGAAGA 103304
```

Qy 550 attgtcaagattgatctagctagctcccttttgagcaagtgtg 590
|| ||||| | | | | | |
Db 103305 TTTCATTCACTGATGCCGGGCAATCCAAAGACTCGTCGT 103

RESULT 8

BSUB0001	BSUB0001	213080 bp	DNA	BCT	26-NOV-1997
LOCUS	Bacillus subtilis complete genome (section 1 of 21): from 1 to				
DEFINITION	213080.				
ACCESSION	Z99104	AL009126			
VERSION	Z99104.1	GI:3632267			

SOURCE *Bacillus subtilis*.

ORGANISM	Bacillus subtilis	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus.

1 (bases 1 to 213080)

REFERENCE
AUTHORS

Kunst, F., Ogasawara, N., Mosser, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Brons, R., Boursier, L., Brans, A., Braun, M., Briqnel, S. C., Bron, S., Brouillet, S., Brusch, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emmerson, P. T., Etian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Halech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinols, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogihara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Prescan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serrero, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Yamakoshi, A., Tanaka, T., Terpstra, P., Tognotti, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedder, E., Wedler, H., Weitzneger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Daechin, A.

TITLE
The complete genome sequence of the gram-positive bacterium
DANCIII, A.

JOURNAL
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)

MEDLINE 98044033

REFERENCE
2 (bases 1 to 213080)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

TITLE
JOURNAL
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

	FEATURES	SOURCE
1.	100% pure	100%
2.	100% pure	100%
3.	100% pure	100%
4.	100% pure	100%
5.	100% pure	100%
6.	100% pure	100%
7.	100% pure	100%
8.	100% pure	100%
9.	100% pure	100%
10.	100% pure	100%
11.	100% pure	100%
12.	100% pure	100%
13.	100% pure	100%
14.	100% pure	100%
15.	100% pure	100%
16.	100% pure	100%
17.	100% pure	100%
18.	100% pure	100%
19.	100% pure	100%
20.	100% pure	100%
21.	100% pure	100%
22.	100% pure	100%
23.	100% pure	100%
24.	100% pure	100%
25.	100% pure	100%
26.	100% pure	100%
27.	100% pure	100%
28.	100% pure	100%
29.	100% pure	100%
30.	100% pure	100%
31.	100% pure	100%
32.	100% pure	100%
33.	100% pure	100%
34.	100% pure	100%
35.	100% pure	100%
36.	100% pure	100%
37.	100% pure	100%
38.	100% pure	100%
39.	100% pure	100%
40.	100% pure	100%
41.	100% pure	100%
42.	100% pure	100%
43.	100% pure	100%
44.	100% pure	100%
45.	100% pure	100%
46.	100% pure	100%
47.	100% pure	100%
48.	100% pure	100%
49.	100% pure	100%
50.	100% pure	100%
51.	100% pure	100%
52.	100% pure	100%
53.	100% pure	100%
54.	100% pure	100%
55.	100% pure	100%
56.	100% pure	100%
57.	100% pure	100%
58.	100% pure	100%
59.	100% pure	100%
60.	100% pure	100%
61.	100% pure	100%
62.	100% pure	100%
63.	100% pure	100%
64.	100% pure	100%
65.	100% pure	100%
66.	100% pure	100%
67.	100% pure	100%
68.	100% pure	100%
69.	100% pure	100%
70.	100% pure	100%
71.	100% pure	100%
72.	100% pure	100%
73.	100% pure	100%
74.	100% pure	100%
75.	100% pure	100%
76.	100% pure	100%
77.	100% pure	100%
78.	100% pure	100%
79.	100% pure	100%
80.	100% pure	100%
81.	100% pure	100%
82.	100% pure	100%
83.	100% pure	100%
84.	100% pure	100%
85.	100% pure	100%
86.	100% pure	100%
87.	100% pure	100%
88.	100% pure	100%
89.	100% pure	100%
90.	100% pure	100%
91.	100% pure	100%
92.	100% pure	100%
93.	100% pure	100%
94.	100% pure	100%
95.	100% pure	100%
96.	100% pure	100%
97.	100% pure	100%
98.	100% pure	100%
99.	100% pure	100%
100.	100% pure	100%

```

FEATURES
source
Location/Qualifiers
1. .213080
/organism="Bacillus
/strain="168"
/dh_vref="rayon.142

```

gene

CDS

```

/gene="dnaA"
/function="initiation of chromosome replication (DNA
synthesis)"
/note="alternate gene name: dnaH, dnaU, dnaK"
/codon_start=1

```

```

/transl_table=11
/protein_id="CAB11777.1"
/db_xref="GI:2632268"
/db_xref="SWISS-PROT:P05648"
/translation="MENILDLNQAALQIAIEKLSKSPFETWMKSTKAHSLQDGLTLTIT
APNEFARDWLESRLYHLIADIITELTGEELSIKFPVIONODVEDFMPKQPVKKAVKED
TSDPQNNLNPKYTFDTYIGSGNRFAHAASVAEAPAKAYNPLFIYGGVGLGKTHL
MHAICHVYIDHNPQAKVYVILSSSKFTNEFINSIRDNKAVDFRNRYRNVDVLIDITQF
LAGKEQTEPEFHTNHLHESKQIVLISSDRPKPEIPTLEDRLSRFENGLITDITPP
DLFWJIAILRKKAKAEGLDIPNEMVLYIAQIDNSIRELEGALIRVAVYSSLLKINDI
ADLAAEAKDLIPSSKPKVITIKETIRVQVQQFNKLEDFKAKKRTKSAVFAPQIAYI
LSREMTDSSLPKIGEEFGRDHTVTIHAHEKISKLLADDEQLQOQHVKEIKEQLK"
1939..3106
/gene="dnaN"
CDS
1939..3075
/gene="dnaN"
/EC_number="2.7.7.7"
/function="DNA synthesis"
/note="alternate gene name: dnaG, dnaK"
/codon_start=1
/transl_table=11
/product="DNA polymerase III (beta subunit)"
/protein_id="CAB11778.1"
/db_xref="GI:2632269"
/db_xref="SWISS-PROT:P05649"
/translation="MKPTIQDRLVESVDVLKAVSSRTTIPILTGIKIVASDDGVSF
TGSDDSISEFPKPEGDEKIYVTIQPGSIVLOARFEIVKKLPMATVEIENVQVY
LTIIRSKAEENLGLDADSEYPLHPQIEEHHAIQITDLLKNIROTVAVSTSETRP
ILTVGNKPEQSELQVLTQASHLRAKLKADIPEDRSYNNVIFPGKSLTELSKILLDDN
OELVDIYITQVLFKAKNYLFFSRLLDGNYPDTTSLIPQDSKTEIIVAKTKFLAID
RASLLAEGRRNVVKLSAKPAESIEISSNSPEIGKVVEAIVADQIEGEELNISFSPKY
MLDAKVLGEGAEIRVSTGAMRPFILTPNDETIVQILLPVRTY"
3087..3106
/terminator
/gene="dnaN"
3206..3421
/gene="yaaA"
3206..3421
/gene="yaaA"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB11779.1"
/db_xref="GI:2632270"
/db_xref="SWISS-PROT:P05650"
/translation="MANPISIDEMITTLQGLKLADVIQSGMAKWFLEHSEVLVND
PDNRRGKLYVGDDVEIEGEGSQVYN"
3437..4549
/gene="recF"
3437..4549
/gene="recF"
/function="DNA repair and genetic recombination"
/codon_start=1
/transl_table=11
/protein_id="CAB11780.1"
/db_xref="GI:2632271"
/db_xref="SWISS-PROT:P05651"
/translation="MYIQNELTSIRNYDHAELQFNKVNVIIGENQAQKYNLMEALY
VLSNAKSHRTSNKELTRDKDYAKTEGRVMQNGAIPMQLVISKKGKGVNHLHQQ
KLQYVGALNTIMFAGPDSQPQRRRLEIDMEIQGSPVYLHDLSLYQKILSQ
NHFLQQLQTKRQTDRTYMLDVLDTQLVQVAAKVVVKKLQVTAOLEKWAQPIHAGISRGL
EELTLKYHTALDVSDDLKSGIDSGEAFSKLRKEIERGVTLCSPHRDDVLYFYNG
RDVQYTSQCQQRRTALSKLAIEDILHIEIGEYPIILLDDVLSELDDYRQSHLHTI
QGRVQTVETVTTTSDVDGIDHETLROAGMFRVQNGALVK"
4567..4725
/gene="yaaB"
4567..4725
/gene="yaaB"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB11781.1"
/db_xref="GI:2632272"

```



```
/sub_species="aureus N315"
/db_xref="taxon:158879"
241..957
/gene="SA0245"
241..957
/gene="SA0245"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB41469.1"
/db_xref="GI:13700171"
/translation="MKYAGILAGGIGSRMGNNVPLPKQFLDLDNKPILHTLEKFIILN
DFEKIIITAPQOMTHTKRIKFKISDERIEVIGGSDRNDTINIVKIHSTNGIN
DDVIVTHDAVRFPLTHRIKENIQAALEYGAVDTVIDAIDTIVTSKDNQIDAIPVR
NEMVQGTPOQSFNNLLKESYAQLSDEQKSILSDACKIIVETNKPVRVLKVELYNIKV
TTPYDLKANAIIRGGIADD"
950..1975
/gene="SA0246"
950..1975
/gene="SA0246"
/note="ORFID:SA0246"
hypothetical protein, similar to D-xylulose reductase"
/codon_start=1
/transl_table=11
/protein_id="BAB41470.1"
/db_xref="GI:13700172"
/translation="MINQYOLVAPROFEVTVNNVDIYSDVIVIRPLYMSTCAADORY
YTSRDENVLSQKLPSMLIHGCVGEVFDKSGVNGTKVVMVNTPTERKDDVLAENY
LKSSYFSSGHDGFMQDFVLLNHRAPVLPDDIDLSIISYTELVTSUHAIRFEKKS
ISNKTGIFMGDGNLGYITAILLRKLYPESKIYVFGKTDYKLSHFSFVDDVFFINKIP
EGLTFDAIEFCVGGQSGQSAIQNMIDYISPEGSIALIGVSEFPVEVNTRLVLEKGLTL
IGSSRSGSKQDFQDVVDLYIQYPDIVDKLALLKQGEYEIATINDLTFEADLSTSWGK
TVLKWM"
1997..3685
/gene="SA0247"
1997..3685
/gene="SA0247"
/note="ORFID:SA0247"
hypothetical protein, similar to teichoic acid
biosynthesis protein B"
/codon_start=1
/transl_table=11
/protein_id="BAB41471.1"
/db_xref="GI:13700173"
/translation="MVKSKIYIDKIYWERVOLFEVGHSENLDLEDNFRNLNLTETRT
MKANDVIDGNOFCVRENVAILDNGYVLPEDKYLVLVNEQELDYIAQLNPQVINDAYON
LKPEEVEYNELETFONGKINFLLOTYLKEPKGSGISKTYVTTPPEISSDNGNEVLQD
VWTPPEVKSIIYIKRYELRKIRFKQSFNTKQFIFKALFNTRFFHLKKNGNVLFTSD
SRPTMSGNFYIYNEMLRQNLDDKYDHTVFKANITDRGIIIDKFRPLPYLLGRADYIF
VDDPHLIYTVFRRSQEVIVQVHAGAFKTVGFSRTGKGGPPIIDSLNHSYTKAYF
SSETDIPFYAEAFGIEKKNVPTGVPRTDVLDEAYATQIKQEMDELPIIKGKKVIL
FAPTFRSGHGHTAHYPPFKIDFERLARCYCKNNAVLFKMHPFVKRNLNADKHQOYE
VDVSDFRVNDLIFITOLLISDYSSLIYEAVFKPMIFYAFDLEDITTRDRFEPYE
SFVPCKIYQSFDALMDALNEDYEGEKVIPFLDKHFYIQGRSSERLVRNLFGS"
3718..5439
/gene="SA0248"
3718..5439
/gene="SA0248"
/note="ORFID:SA0248"
hypothetical protein, similar to beta-glycosyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAB41472.1"
/db_xref="GI:13700174"
/translation="MKKFSVIVPTYNSEKVIYITELLNSLAKQDFPKTEFEVVVDDCST
DQTQIVERTKRNKLVNQSLTEINSGGPKRPNVALKQAGEFVFLVDSDDYINKETLT
KRAAAFIDEHSDDLKMKGVNGRGPQSMFKETAPEVTLNRSRIITYSPTKIYRT
ALLKNDNDIYFPEELKSAEDQLFTMKAYLNANRISLVRNSDKAYIYATKREGHMSAYVS
PEDFYWMRLIAVLEIADLEEAHKDQILAEFLNRHFSFRTNGFSLKVKLEDPQWI
NALGDFTOAVPERDALVMSKLRPLHLHARAKDIDNRYTVEESYRGQGYRFQDIDCK
LNIOFNEGSEYFEGIDIAKPKVMTAFKFDNHKIYVTELTLINEPMIGEGHYDVRKLHNS
RNKHTHYVLSLVNANKQYRFNIMDLKAYLPKEIKWDVDFLEVOIGTEVFEVRVGNQ
```

```

RKNYAYTAETSAIHLNNDFYRLTPYFTKDFNNISLYFTAITLTDSDISMKLKGNKII
LTGLDRGYVFEEGMASVVLKDDMIMGMLSQTSENEVEILLSKDIKKRDFKNIVKLNTA
HMTYSLK"
5583..6257
/gene="scda"
5583..6257
/note="ORFID:SA0249"
/codon_start=1
/transl_table=11
/product="cell division and morphogenesis-related protein"
/protein_id="BAB41473.1"
/db_xref="GI:13700175"
/translation="MINKNDIVADIVIDYPKAADIFRSVIGDFCGGQGVSTEAASLEK
KNVDLNEQLRDNDEQNTNTPGSLNPKFLNVSLSIQYIOAAVHBPRLREFKNLTPTVY
KLSKVHPGNHPLYVELKETVDFKSGMLEHMQEDDDVDFPKLIKEYOGEVVDNNTVI
DDLVSDIATGQLLVKMSDLTSSYEPPIEACGTWRLVYQRLKALEVLTHEHVHLENHV
LFRKVS"
6502..8256
/gene="lyts"
6502..8256
/gene="lyts"
/note="ORFID:SA0250"
/codon_start=1
/transl_table=11
/product="two-component sensor histidine kinase"
/protein_id="BAB41474.1"
/db_xref="GI:13700176"
/translation="MLSLTMLLERVGLIILAYVLMNIPYFKNLMNRRRTWKARWQL
CITFSLALMSNLGTIVIDHOHSLSGSVYFRLLDDVSLANTRVLITIGVAGLVGGPFVG
LFVGVISGIFRVYMGADAQVYLISIFIGIAGYFGLQQRKRYPSTIAKSAMIGIV
MEMQMSILITFSDHKAYAVDLISIALPMLIIVNSVGTAFPMSTIIISTLKQEQMKAV
QTHDVLQMLQTLPPYFKEGLNRESAQOIAMIKNLMKVSVAITSKNELLSHVCGAGSD
HHIPTNELILYSKDKVLKSGKLEKVEHTKEEGCSHPNCPPLAAIVIPLEMHGSIYGTIL
KMFTNPNDLTFVQRLAEGLANIFSSQIEGAEATOSKLLDAIKSLAQVSPHFF
FNSINTISALVRINSEKARELLELSYFFRANLQSGQHTITLDEKLSQVRAYLSLEQ
ARYPGRENINVEDKYRDLVLPPLIQLIIVENAIKHAFTNRKQGNIDSVSKETAT
HVRIIVQDQGGISKDKMHLIGETSVESSEGTSALENLNLRLKGLFGKSAALQFEST
SSGTFWCVLPLYERQEEE"
8259..8999
/gene="lytr"
8259..8999
/gene="lytr"
/note="ORFID:SA0251"
/codon_start=1
/transl_table=11
/product="two-component response regulator"
/protein_id="BAB41475.1"
/db_xref="GI:13700177"
/translation="MKALIIDDEPLARNELTYLLNEIGGFEETNEAENVKETLEALLI
NOYDIIFDLVNLMDENGIELGAKIQKMKPEPAIFATAHQYAVQAPELNATDYILKP
FGOKRIEQAVNKVRATKAKDDNNASATANDMSANFDOSLPVEIDDKTHMLKQONIIGI
GTNGITPTIHTNHKIYETPLNRYEKRLADTYTIRHSRVIIINTKHKEVQOQFNFT
YMWILTVGNKMQVGRSPMKDFKASIGLL"
9118..9555
/gene="lrgA"
9118..9555
/gene="lrgA"
/note="ORFID:SA0252"
/codon_start=1
/transl_table=11
/product="holin-like protein LrgA"
/protein_id="BAB41476.1"
/db_xref="GI:13700178"
/translation="MKQKDASKPAHFHQHVIVITALVLFVSKIIIESFMPIMPASVIG
LVLLFVLCTGAVKLGVEKVGTTLTNNIGLLFVPAGISVNSLGVISQAPFLIIGLI
IVSTILLICTGYVTQIIMKVTSRKSGDKVTKIKIEAAQH"
9548..10249
/gene="lrgB"
9548..10249
/gene="lrgB"
/note="ORFID:SA0253"
/codon_start=1
```



```
/transl_table=11
/product="holin-like protein LrgB"
/protein_id="BAB41477.1"
/db_xref="GI:13700179"
/transl_table=11
translation="MINHLALNTPYFGILLSVIPFFLATILPEKTRFFLEFAPLFVSM
VFGVAFLYLTGPKYKIGGDIIEFLEPQATFAIPLYKKEVLVKKHRRIGIG
IGTVALLILTFAPKLAQFANDVILSMQATATALPVSAGIGGKEILTSIAVLIN
GVIIYALGNKFLKLRITNPDIANGALGTSGHTLGVAPAGGIVVEESMASIALVLVG
VVVAVVPVFVAFVFF"
complement(10357..11061)
/gene="SA0254"
CDS
complement(10357..11061)
/gene="SA0254"
/notes="ORFID:SA0254
hypothetical protein, similar to transcription regulator
GntR family"
/codon_start=1

Query Match      20.5%; Score 131; DB 1; Length 301050;
Best Local Similarity 55.2%; Pred. No. 7.4e-26;
Matches 323; Conservative 0; Mismatches 250; Indels 12; Gaps 3;

QY 15 tttagtctctttaggaccagaggagcagcagcaagaccagtgcttttagagcgtctgct 74
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215548 TTTTATACCTTTGAGGCCCGCAGGCTCTGGAATAACACTGTAATTAATGAAGTTA 215607

QY 75 accaattttagagaaaaagtagagggtgtgacgacctgacacctgacgagctt 134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215608 CCATAGATTA--GTAAGAATTATGATGTCATTATGACATAGAACACCGTGTGTTCC 215664

QY 135 gattggggagaagattcgggaagtatttgggtcccaagtcatactcagatggatgtaa 194
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215665 TACTGTCGAGAATACTGTAATAATTGATTAGAGCAATGAT-----ATGGACATTAG 215718

QY 195 aacagagctactctctatattgacagtcagacacacatttgggaaaaaagtcttc 254
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215719 AACTGAAGCAATGTTATTTCGTCGATCTGAAGAGAACATCTTGTATTAAGGTCATACC 215778

QY 255 agccttgaagctggcgaagttggtcatcattggtctttatcatgattctgttgccta 314
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215779 AGCTTTAAAGAAAGTAAGTTGTTGTGTCGTCGTATATCGATAGTTCATTAGTTA 215838

QY 315 tcaggggatttgctggttagatattgaagccattgactggctcattcagtttgcgac 374
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215839 TCAAGGTTATGCTAGAGGGATTGCGTGTGAAGAGTAAGAGCATTAACGAATTTGCAAT 215898

QY 375 agatgggctcaaacccgatttgacactctatttgacatcgagtggaagaagggctggc 434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215899 AAATGGATTATATCCAGACTTGACGATTTATTTAAATTTAGTCTGAAGTAGTCGCGA 215958

QY 435 tcgtattgctgcaatagtagccgcgaggttaactcgttggattggaaggggttgacct 494
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215959 ACGTATTATT--AAATTCAGAGATCAAAATAGATTAGATCAAGAAATTTAAAGTT 216015

QY 495 gcataaaagtttgctcaaggtcaccttctctctctggtgataagagggaatcgattgt 554
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216016 TCACGAAAAAGTAATTGAAGTTACCAAGAAATCATTCATAATGAATCAACACGGTTCAA 216075

QY 555 caagattgactgtagtctcccttggagcgaagttgtggaactac 599
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216076 AAGCGTTAATGCAGATCAACCTCTTGAATAATGTTGTTGAAGACAC 216120

RESULT 10
AP003359 AP003359 343590 bp DNA BCT 30-MAY-2001
LOCUS Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
DEFINITION sequence, section 2/9.
ACCESSION AP003359 BA000017
VERSION AP003359.2 GI:14246074
KEYWORDS Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
SOURCE strain:Mu50) DNA.
```

ORGANISM

Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.

REFERENCE

AUTHORS

Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iian,J., Ico,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Ui,Y., Kobayashi,N., Tanaka,T., Sawano,T., Inoue,R.,
Kaito,C., Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S.,
Yabuzaki,J., Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K.,
Yoshino,C., Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and
Hiramatsu,K.

TITLE

Whole genome sequencing of meticillin-resistant Staphylococcus

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@sakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)

COMMENT

FEATURES

source

Location/Qualifiers
1..343590
/organism="Staphylococcus aureus subsp. aureus Mu50"

/strain="Mu50"

/sub_species="aureus Mu50"

/db_xref="taxon:158878"

441..1331

/gene="SAV0307"

441..1331

/gene="SAV0307"

/note="similar to outer membrane protein precursor"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB56469.1"

/db_xref="GI:14246075"

/translation="MNKISKYIAIASLVAVTSAPOTTNSTAFKSSAEVQQTQAS
IPASQANLGNINIMAVYQNSAEAKALYQYNSAKTQLDKEIKKNGKHKLIAIAL
DLDETVDLNSPYQGYASIHNNPPEGWQAAKAPVYGAKEFLYADKGVDIYV
ISDRDEKDLKATQKNLKOQIPQAKKSHLLKQDDKSKESBROMYQDKHLMLEFG
DNLLDTDPKEATAESREALIEKHKDDFGKKYIIFPNPMTGSEWATYNNNYKASDKA
KKLRKNAINKQFDPKTEV"

1580..2629

/gene="SAV0308"

1580..2629

/gene="SAV0308"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB56470.1"

/db_xref="GI:14246076"

/translation="MFLANNEIRNKLKGLIIGVLTMIISVLLFLLSLGLINMKNK
EGIDKQDAIVLKNQANQVQSVFNKKDIENKQKQATLKOTGEIVSNGHOKDNVL
VFGVEKSSFLVPSLIEGHKATKDNELADTKNKGFKIGDTLSLSQDEKLHIVGFT
ESAKYNASPVITNDATIAKINPLRTGDKINAVVVRDTRNNKDKLNQDELAIVNDFI
ENLPYKQPQNLNFMISFLVISATVIGFLVYMTLQKTSFLGILKAQCGFTNGYLAN
VVISOTLILALFCTAFGLLLLTGVTGAFDPVAPVKPFDVLTLLVFAIVLMIVSVLSLFL
SILTIRKIDPLKAIG"

2642..3319

/gene="SAV0309"

2642..3319

/gene="SAV0309"

/note="similar to ABC transporter ATP-binding protein"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB56471.1"

/db_xref="GI:14246077"

RESULT 10

AP003359

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

```
/translation="MLKPENVTKSKFGDNRNIEAVKDTNFEINKGDIALLVGPSSGSK
STFTMAGATQPTSGHLLINNODITTMKOKALAKVRMSEIGFLLQATNLVPFLTVKQ
QFTLLKKKNVMNEDYQQLMSGLTSLNKLPSSETSGGKORVAIAKALYTNPSI
ILADEPTAALDTEANAIEVIKILRQAKRKKACIIVTHDERLKAICYDRSYHMKDGVLEN
LENETVE"
3527..4558
/gene="SAV0310"
CDS
3527..4558
/gene="SAV0310"
/note="similar to regulatory protein Pfor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56472.1"
/db_xref="GI:14246078"
/translation="MDLMIGTFLFLVLVITFLFTYKAPNMRMAGALANAIAIATFLV
EAPNKYVGGEVFGKFLLEELDAGGLGGVAAGLTALAGSPVYALVIAAAGGMD
LLPFEPAGYIGYVNMKYTEKYVPGCVBLIGSIVILAPLARLIAVLLPVVNSTLRIG
DIQSSNTNPDIIMGIIILGGIITVVGTAPLSSMALTALLGLTGVPMAIGAMAFFSAP
MNGTLFHLRLKGDKRSTIAVSIIEPLSQADIVSANPIIYITNFFGGAIGLIAMSGL
INDATGTATPIAGFLVMFGFNHPMTIIVYGVMAIVGALAGLYGSIVFKKYPVIVTKQD
MINRGAVDA"
4901..6019
/gene="SAV0311"
CDS
4901..6019
/gene="SAV0311"
/note="similar to carboxydrate kinase PfkB family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56473.1"
/db_xref="GI:14246079"
/translation="MSDSEKILKRIKDNPPFISORELAEALGSRPSVANIISGLIOK
EYVNGKAVLNEDYPIVCIGAANYDRKFYVHKDIAVETSNPTVSTSTRISGGVARNIAEN
YGRGETVAFLSAGSDSEWENIKRLSTPPMNLDRVQOFENASTGSTALISKEDMT
YGLADMEVDFITFELFKRSHLKKAKCIIVDNLKREALNFCAYTTHQIKLVIT
TVSSPKMNPDSLSHAIDWIITNKRDETEYINLKIESTDDLKIAAKRWNDLGVKNVIV
TNGKELIYRSGEEIISKVMPNSVRKVTGAGDSFCAAVVYSLWNGMSTEDILIAGH
VNAKKTETRYTRQNLDDQQLYHDMEDYKNGKTKVY"
5994..6917
/gene="SAV0312"
CDS
5994..6917
/gene="SAV0312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB56474.1"
/db_xref="GI:14246080"
/translation="MANLQKYEYSREVOQARENQPIVALESTIISHGMPYPQNVEM
ATTVEQIIRNGATIPATIAIDGKIKIGLESEDLILATSKDVAKVRRLDAEYVAMK
CVGATTVATTMICAMAGIOFFVTGGIGGVHKGAEHTMDISADLEELSKTNVTVICAG
AKSLDLPKTMEYLETGKVPVIGVOTNELPAFTFRESGVKLTSSVETPERLADHLTK
QQLNLEGSIIVANPIPEYHALSKAYIEAINEAVEAENQGIKGDATPFLLGKIVEK
TNGKSLAANKLVENNAALGAKIAVAVNKL"
6928..8148
/gene="SAV0313"
CDS
6928..8148
/gene="SAV0313"
/codon_start=1
/transl_table=11
/product="probable pyrimidine nucleoside transport
protein"
/protein_id="BAB56475.1"
/db_xref="GI:14246081"
/translation="MNILFALTGIATFALFVAFLESFDRKNIDFKKLTIMIFIQVLIVL
FMNMTTGLITLTAFFFEGLINVSAGINFVGDIONKNGFTFFLNLLPLVFISV
LIGFNTIKVLPFIKYVTAINKITRMGRLESYFAISTAMFQGPVEYLTIKDIPLR
SRAKLYTIATSGSAVSMAMLGSMQMEPKFVYTAVMLNIFSALITASVINPKPSDD
SDVEIDLTKSTETKSVNGTKGPKKVAFFQMIGDSAMDGFKIAVAVVALLAFISM
EAINMFGSVGLNFKLIGIVFAPIAFLMGPWPSEAVPAGSLMAKTLITNEFVAMLOF
KNVLGDSARTQGLIISVLSFANFVTGVIIVGSIKIGISDRQGEKVSFAMRLLGST
complement(8253..9785)
```

```
/gene="SAV0314"
complement(8253..9785)
/gene="SAV0314"
/note="similar to sodium-coupled permease"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56476.1"
/db_xref="GI:14246082"
/translation="MKBVGFGLNWAIVVILLAMLFIVGVETKRASQSTNSPFTASG
RLPSWVVGFSIYATLTSJAITFMSPTPEKAFLTDSYIAGNIAIVAIIPLLIVFYVPEFK
KLKTSAYEYLEARFGPSIRVIGSLLFVWHGLRVAIVIYLPDLTAISVSDMNPYIVA
SLVGLLCILYTFGLGFGVGVSDFIQGVLLGGALVILIGVMNKGFGFVFADAIIE
HKLISADNNKLTAAAGAIPIIFLGNFNLLYQYTSQDVVQRYQASDSLKETNKSLLW
TNGLIALISAPLFYGMGTMLYSFYTHEAVLPKGFNTSSVVPYFILTMPPPFVAGLLIA
AIPAAASNTSSLSNISCISIDIKORFEGKDSERHEVNFARFIIITAGIFFGMSL
YLTAASNDLMDLFLFVTLGFLVPLAGVAVGIFTKRTNTFGVIGLILGIIIFAVVN
GVGKGNPFVYVSTISFVAVFAYILSFIVPSKKHKDITGLTIFEKDKPSTYISKATAT
NK"
complement(9825..10706)
/gene="nanA"
CDS
complement(9825..10706)
/gene="nanA"
/note="SAV0315"
/codon_start=1
/transl_table=11
/product="N-acetylneuraminate lyase subunit"
/protein_id="BAB56477.1"
/db_xref="GI:14246083"
/translation="MNKDLGLYAALLVPDENQVNEQGLKQIAQNAITEELDGLY
VNGSSGENFLNTEQKQVFKVAEYGDVKLQAQVSLDNLNIAIEQLKYATELGYD
ALGSAVTPFYPTFEERLYYFDIEATQNNMIYIIPDLTGVNISIEQFSELFNHEK
IVGVKYTPAFNELLERKAPDPKDLILSGDEMLVQATISGVDGAIGSTLVNNGRRAR
KIFDLARQGOIQEAYOLQHDSDNIIETVLSMGIYPTLKEILRHRDIDAGLKRPEKPF
NEARQRLDQLIAKYDL"
10864..11724
/gene="SAV0316"
CDS
10864..11724
/gene="SAV0316"
/note="similar to glucokinase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56478.1"
/db_xref="GI:14246084"
/translation="MYVIAIDIGGTQIKSAVIDQLNMFDYQQIISTPDNKSELITDKV
Query Match 20.5%; Score 131; DB 1; Length 343590;
Best Local Similarity 55.2%; Pred. No. 7.5e-26;
Matches 323; Conservative 0; Mismatches 250; Indels 12; Gaps 3;
Qy 15 tttagtctcttgaggagaccagaggagcagcagaccagtgcttttagaggctcgtcgt 74
Db 187808 TTTTATAACTTTTGGGCCCCAGAGGCTCTGGAAGAACAACTGTAATAATGAAGTTTA 187867
Qy 75 accaattttagagaaagaggagtagaggtgttgacgacccgtgaacctggcgagctctt 134
Db 187868 CCATAGATTA---GTAAAAGATTATGATGTCATTTGACTAGAACACAGGTGGTTC 187924
Qy 135 gattggggagaaagattcgggaagtgtatttggatccaagtcatactcagatggatgctaa 194
Db 187925 TACTGGTGAAGAAATACGTAATAATTGTTATTAGAAGCAATGAT-----ATGCACATTAG 187978
Qy 195 aacgagagctactctctctattatgccactcgcagacagcatttggkggaaaaagttctcc 254
Db 187979 AACTGAACCAATGTTATTTCGTGCATCTAGAGAGAACAATCTGTATTAAAGGTCATACC 188038
Qy 255 agcccttgaagctggcaagtggcatcatgatgatctgtttatcagatgtcttctgtgccta 314
Db 188039 AGCTTTAAAGAGAGGTAAGGTTGGTTGTGTGATCGCTATATCGATAGTTTCAATTAGCTTA 188098
Qy 315 tcagggtatttgctggtgcttagatatttgaagccattgactggctcaatcagtttgcgac 374
```

D	b	188099	TC	AAGGTTATCGTAGAGGATTGGCGTTGAAGAAGTAAAGACATTAACAGCAATTTCGCAAT	188158
QY		375	agatgggctcaaacccgatttgacactctctttagacatcgagggtgaagaaggcgtgc	434	
D	b	188159	AAATGGATTATATCCAGACTTGACGATTATTTAAATGTAGTCTGAAGTAGTCCGGA	188218	
QY		435	tcatattgctgcgaatagtaccgcgaggttaactcgtttggtattggaagggttgacct	494	
D	b	188219	ACGTATATTT---AAAAATTC	AAGAGATCAAATAAGATTAGATCAAGAAGATTTAAAGTT	188275
QY		495	gcataaaaagtctgccaagcttaccttcttcctgtagataaaagaggaaatcgactgt	554	
D	b	188276	TCACGAAAAGTAATTGAAGGTTACCAGAAATCATTCATTAATGAATCACAACGGTCAA	188335	
QY		555	caagattgatgctagctccctcttgagcaagttgtgaaactac	599	
D	b	188336	AAGCGTTAATGCAGATCAACCTCTTGAAATCTTGTGAAGACAC	188380	
RESULT		11			
LOCUS		MC109	971 bp	DNA	BCT 18-AUG-1995
DEFINITION			M.Capricolum DNA for CONTIG MC109.		
VERSION		Z33079			
KEYWORDS		Z33079.1	GI:541703		
SOURCE			atp-binding pyrimidine kinase.		
ORGANISM			Mycoplasma capricolum.		
REFERENCE			Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;		
AUTHORS			Entomoplasmataceae.		
TITLE			1 (bases 1 to 971)		
JOURNAL			Gillevet,P.		
REFERENCE			Bork,P., Ouzounis,C., Casari,G., Sander,C., Dolan,M. and		
AUTHORS			Gillevet,P.		
TITLE			Exploring the Mycoplasma capricolum genome: A Parasite reveals it's		
JOURNAL			Physiology		
REFERENCE			Unpublished		
AUTHORS			2 (bases 1 to 971)		
TITLE			Gillevet,P., Ally,A., Barton,F., Brenner,S.E., Clark-Whitehead,R.,		
JOURNAL			Dolan,M., Douglas,N., Hsu,E., Purzycki,M.S., Richter,B., Russo,S.,		
REFERENCE			Sartell,J., Smith,S.W., Wang,C., Williams,J. and Gilbert,W.		
AUTHORS			Direct Submission		
TITLE			Submitted (15-SEP-1994) Mycoplasma capricolum Genome Project,		
JOURNAL			Harvard University, Cambridge, MA 02138 Prokaryotic Genomes Project		
REFERENCE			Institute for Computational Sciences and Informatics George Mason		
AUTHORS			University, Fairfax, Virginia, USA, 22030-4444 E-mail:		
TITLE			gillevet@uranus.ncchr.nih.gov		
JOURNAL			3 (bases 1 to 971)		
REFERENCE			Bork,P., Ouzounis,C., Casari,G., Schneider,R., Sander,C., Dolan,M.,		
AUTHORS			Gilbert,W. and Gillevet,P.M.		
TITLE			Exploring the Mycoplasma capricolum genome: a minimal cell reveals		
JOURNAL			its physiology		
MEDLINE			Mol. Microbiol. 16 (5), 955-967 (1995)		
COMMENT			6059641		
			NOTES:		
			The tga codon is read as Tryptophan in Mycoplasma capricolum Coding		
			sequences below were putatively identified by Peer Bork using the		
			program Genequiz at the EMBL.		
			EMBL, Meyerhofstr.1, 69012 Heidelberg, Germany		
			E-mail bork@embl-heidelberg.de		
			More than 870,477 bases have been sequenced by the genomic walking		
			technique and assembled into 372 non-overlapping contigs (Accession		
			numbers Z33005 to Z33376) covering 214,528 base pairs. The length		
			of the contigs varies from 63 base pairs to 2049 base pairs with an		
			average length of 658 base pairs. The data consists of 13,091 bases		
			(6%) from one pass fluorescent sequencing and 201,437 bases (94%)		
			produced by genomic walking. Of the latter, 155,805 bases (77%)		
			have multiple coverage on at least one strand and 112,621 bases		
			(56%) are covered on both strands. There is a total of 968		
			ambiguous nucleotide assignments in the data set.		
			0.5%). We have compared 8,868 bases of our data with Mycoplasma		
			capricolum sequences already stored in the public databases and		
			note less than 0.7% difference between the two data sets (
			including ambiguous calls, insertions, deletions and mismatches).		
FEATURES					
source					
CDS					
Location/Qualifiers					
1..971					
/organism="Mycoplasma capricolum"					
/strain="ATCC 27343(Kid)"					
/db_xref="taxon:2095"					
84..380					
/partial					
/note="ORF identified by homology to Swissprot entry					
gp 219544 ASFVA240L_1; KTHY_HUMAN; Probability 5.2-05"					
/codon_start=1					
/transl_table=4					
/codon=(seq:"tga",aa:Trp)					
/evidence=not_experimental					
/product="ATP-bind. pyrimidine kinase"					
/protein_id="CAA83746.1"					
/db_xref="GI:950071"					
/db_xref="SPTRMBL:Q49001"					
/translation="MFITFGMDSGKTTALLKVKEELRLNYKVLITRPPGGEVTA					

Db	4735	CGCCGGCGCTGGCGCGCGCGCGCTGGCTCGATCGCTTACCGATGCCACCTAT	4676
QY	310	gactatcaaggatttgctgctgtagattgaagccattgactgactcattcagttt	369
Db	4675	GCTACACAGGCGGGCGCGCGCTTTCGCGAGGCGCGATCGCGCGCTGGAAGCTTC	4616
QY	370	gcgcagcagtggtctcaaacccgatttgacactctatttgacatgagtggaagaagg	429
Db	4615	GTCCAGGGGACCTCGGGCGCCGACCTCAGCGTGTTCGACCTGCCGCTGGAATCGGC	4556
QY	430	ctggctcgctattgctgcgcaatagtacccgca	461
Db	4555	CTGGCGGGCGCGCGCGCTGCGCTCGA	4524
RESULT	14		
LOCUS	AE001292/c		
DEFINITION	Chlamydia trachomatis section 19 of 87 of the complete genome.		
ACCESSION	AE001292	AE001273	
VERSION	AE001292.1	GI:3328586	
KEYWORDS			
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis		
REFERENCE			
AUTHORS	Stephens,R.S., Kallman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.		
TITLE	Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis		
JOURNAL	Science	282 (5389), 754-759 (1998)	
MEDLINE	99000809		
PUBMED	9784136		
REFERENCE			
AUTHORS	Kallman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hymen,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.		
TITLE	Comparative genomes of Chlamydia pneumoniae and C. trachomatis		
JOURNAL	Nat. Genet.	21 (4), 385-389 (1999)	
MEDLINE	99206606		
PUBMED	10192388		
REFERENCE			
AUTHORS	Stephens,R.S., Kallman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-MAY-1998) Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA		
FEATURES			
source	Location/Qualifiers		
	1. .12333		
	/organism="Chlamydia trachomatis"		
	/strain="D/UW-3/CX"		
	/db_xref="taxon:813"		
gene	216..980		
	/gene="kdsB"		
	/note="CT182"		
CDS	216..980		
	/gene="kdsB"		
	/codon_start=1		
	/transl_table=11		
	/product="CMP-2-keto-3-deoxyoctulosonic acid synthetase"		
	/protein_id="AAC67774.1"		
	/db_xref="GI:3328587"		
	/translation="MFAFLTKVGVILPSRWSSRRPGKPLAKLTKTLVQSYENALSSQSLDCVVATDDQRIFDHVHVEFGGLCYMTSANGTERVEVSRHPQAEIVNV IQGDECLSTPVDIGLVSTLENNPAADMTVSTETDPEALLTDHKVKVCVFDKNGKAL YFSRAIPHNFKPTPIYLHIGYAFKAFLSYKIPPSLSLAEDLEQLVLEIGR SIYHVVNATGPGSDVPDITKVEQYLILCSKASF"		
gene	956..2575		
	/gene="pyrG"		
	/note="CT183"		
CDS	956..2575		

/gene="pyrG"			
/codon_start=1			
/transl_table=11			
/product="CTP Synthetase"			
/protein_id="AAC67775.1"			
/db_xref="GI:3328588"			
/translation="MSPKSIPLTGGVYVSSLGKLTAAASLALLLERQDLKVMKLDPY LNVPGTMNPEHGEVYTDGQVETDLDLGHYHRFSSVOLSKYSTATSGQIYTKVLTK ERNGEFGSTVQVPHVTNEIINVIQSCADHHPDILIVEIGGTIGDTESLPFLAVR OFRGEHQDCLSIHMTVYPLRAAKEIKTPTQHSVQNLRGISGIPDVLGRSEAPLS TEVKRKISLFCNVPEHAFVNAIDLSRIYEMPLLLAKENISDFLLNKGFSKPLDLS DMQDLVEALCDKQRHVRIGLVKGYLEHKDAYKVFESLHFASHPANCSELVETAPE SEDLLEQLSQDCGLIPGGFGTSTRWEGKLSAARCRERNIPFCGICLGMQALVVEYAR NVLDKPLANSMEINPETPDVPMWEGQDSVVKGGTMRLGAYPCRIAPGSLASAAKTT DLVQERHRRHYEVNPSYIERLEHGLKIAGVPLGELCEIVEIPNHRMMLGVQFHPF LSKLAKPHPLFIEETRAAKAYSLEKANHEHR"			
2562..3008			
/gene="yqgF"			
/note="CT184"			
2562..3008			
/gene="yqgF"			
/codon_start=1			
/transl_table=11			
/product="YqgF Family"			
/protein_id="AAC67776.1"			
/db_xref="GI:3328589"			
/translation="MNTAKQQQAFGLIDYKKRIGLAFASSPLLIPLPIGVNEARSSL TLTAAALVSIIRKRAVTVVFGNPLPMQKAYASSVQSEIQELAAALIQEMTAIEVLMD ERLSSAQAERMLKSDCGLNKRQKNRPNPSDLAATLILSFLDSRLY"			
3329..4648			
/gene="zwf"			
/note="CT185"			
3329..4648			
/gene="zwf"			
/codon_start=1			
/transl_table=11			
/product="Glucose-6-P Dehydrogenase"			
/protein_id="AAC67777.1"			
/db_xref="GI:3328590"			
/translation="MKLAVQHFSSSEIDIRVWSELENRIIFYHGAOFSDAEGYSVLKA YLEQLDOQYCTQGNRLFYLSPTPPDYFQEIIRNLNRHOLFVHEQGAQAPWSRLIIEKFP GVNLTAELQCCIDANIDEESVTRIDHLYGKQTVQNLITFANTLFPESCWNQSIYID HVQISVESIGISGRNFFKSGMLRDMQVNHQLTQLCLLTMPEPPSEFFSEETKKEKI KILKILPIREDAVQYQGEVIGVYGLREENVDPNSSVETVALKLFIDNPR WKGVFYLQAGKRLPKRTDIAVIFKSSYNLFAENCPLPLENDLLIIRIOPDEGV ALQFNCKVPGTNILVRPVKMDERYDSYENTVTPAYERLLDCITLGDRTLTFTSNEEVL ASWELSPILLEKWSQVRPIFPNTYAGSLRQEADELISRDGKAWRPY"			
4673..5443			
/gene="devB"			
/note="CT186"			
4673..5443			
/gene="devB"			
/codon_start=1			
/transl_table=11			
/product="Glucose-6-P Dehydrogenase (DevB family)"			
/protein_id="AAC67778.1"			
/db_xref="GI:3328591"			
/translation="MATLISLNDANRMLIADSOEEFLQIACYDWISTANKAIHKRGAF YVALSGKTPLOIQFQEIIVKRAAISDCSKIYVFWGDERANEDVEAGSNYIKAMDILKG LRIPEQDIFRMDADPKGEAYEALIQYVYDAIFDMVMGLGVGDGHTLSLFPETHAL EERKERVFNVEPQLTRMTLTFTPIVQRAHLVAYVQGENKQDLFKHLVHPLGRDTF PIRVGTPLNPQVWLSSDSCRKTDLADIADCKLEMF"			
complement(5440..6312)			
/gene="dnaX_1"			
/note="CT187"			
complement(5440..6312)			
/gene="dnaX_1"			
/codon_start=1			
/transl_table=11			
/product="DNA Pol III Gamma and Tau"			
/protein_id="AAC67779.1"			
/db_xref="GI:3328592"			
/translation="MESAAWEALIQVRQOKLPISAIIILHGQDLSNLSARAYEASLI			


```
/product="conserved hypothetical protein"
/protein_id="AAF39304.1"
/db_xref="GI:7190496"
/translation="MDKDLLENVYRFRYRFFKLSILPALLGLWLFPTDILNVLDTAT
VILSDRVCGILLILLSALSPNPIILRLGVFIGLWVSFSCSPNFSPLVPAHDSLLGF
ATLAIICLLENPREDLEVGTPIETSDYNPSSGKGKAVLLFSLWQSRYLTSIAL
NIADADTCSLTFESSILMWIYSMLIVLSLTGGGERHTRPKAVCVATFMFLFALGATL
IAILLQSLSUTNAGINLTAPVSLAFYDEIRATWOYLTFQSSDRKKLTKVAFYGS
FYRESFWEERSVLPFSKACKQAFGLSPNLVAFACILAIAGFVQINTNLAINPTCR
FVYNACWFTLVLISFSAKSLRHLRWLSLLFAAGILLSPVIFHLPLEAKPLLSIVAS
GIAGIGLSIGRL"
complement(1350..1985)
/gene="TC0451"
CDS
complement(1350..1985)
/gene="TC0451"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39305.1"
/db_xref="GI:7190497"
/translation="MIPLOKLEQAARSFYDQELMLSRONKLSLQDEIHKHKIKSLPI
IFPSALMTGALCALCIGILCFINDLFDYEVFLFELPGILSLAFTALLLFWANKE
QNLVSOKLOQVATCYFESLALCSCPEPKLSVKRLVEFTQDEVLPFGFSKRFFAVL
TLAKPSLLAKESSTKTPPEDEIIEKAFSHIREGLYLSGSKLDHDSQLNQN"
complement(1988..2680)
/gene="TC0452"
CDS
complement(1988..2680)
/gene="TC0452"
/notes="ABC transporter, ATP-binding protein; identified by
match to PFAM protein family HMM PF00005"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAF73555.1"
/db_xref="GI:8163221"
/translation="MLVVKSLCYAHASHFVEDAAAFAPQISFTILGGSGKTTLF
RIIAGLTSTFISGLVWNGQPVQELVAYMOQKEALLPWRVTKNLIILPTLGLPHKQKP
LVQDNFHKVQISGLSSLDLRFDELSSGORQVVFAMOCLEPKPKILLIDEPFTSLD
SLTKEMLYODVRLEAEEGKAILASHVDQCLEIGETTFPIRNRKLLHPLISLQKQGI
AGLLQIKNHLI"
complement(2708..2781)
/gene="trna-ile-1"
/product="trna-ile"
complement(2708..2781)
/gene="trna-ile-1"
complement(2786..2858)
/gene="trna-ala-1"
/product="trna-ala"
complement(2786..2858)
/gene="trna-ala-1"
complement(3049..3777)
/gene="TC0453"
CDS
complement(3049..3777)
/gene="TC0453"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39306.1"
/db_xref="GI:7190498"
/translation="MVFIVIMLSKFYKLSLAVLLINMFSPVVFSAEGTTFIGRMK
SWLSGKTSIPTEDLOTSAITKVSOLLKSKRYDYTGEGSFVQFPFPHSHQIWEI
PQSDLAIRDYTYVAETPNDSYTVVSWVEYPEKVDISKPELNQEGFSGLMILAPESQ
VLYLKAEVOGHKALEFWACEDVDYFGRMLVSVNHTLYQVFMVYKGRSPEVLDEKNI
FVTSFKVTYKVRNKKMDIRKRVSL"
4045..4872
/gene="TC0454"
CDS
4045..4872
/gene="TC0454"
/notes="similar to PID:557478 GB:AE001273; identified by
```

```
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="3-deoxy-manno-octulosonate cytidyltransferase"
/protein_id="AAF39307.1"
/db_xref="GI:7190499"
/translation="MCDYDNLNRFGRVYLRGCDVFAFLTSKRVGILPARMGSSRPP
GKPLAKILGTLKIRSYENALTSRLDCVVVATDDQRIFDHVVVEFGGLCVMTCESCAN
GTRERVAEAVSQPQAEIIVVNIQGDCEPLSPSITDGLVMELEGNPAIDATPVTETVD
PDAILLNHHKVCVDFDTGRALYFSSRVPINNFKRSTPIYLHIGVAFRRFAFLSEYVKI
PPSSLASLAEDLEQLRVLESGRSIYVHVQVQATGSPSDYDPEDISKVEQVLLCPSKVSF"
4848..6458
/gene="TC0455"
CDS
4848..6458
/gene="TC0455"
/notes="similar to PID:557479 SP:O59321 GB:AE001273;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="CTP synthase"
/protein_id="AAF39308.1"
/db_xref="GI:7190500"
/translation="MSFKSIFLTGGVYSSLGKGLTRASLALLLERQGLKVAMLKLDPY
LVNDPGTMNPEYHEVYVTDGVEDLGLDGLHFRFSVSQKSKSTATSGQIYTRVITK
ERNGEFLGTSVQVIVHTNEIINVIQACAEHHRKPDVLIIEIGTIGDIESLPLEAVR
QFRCHEPQDCLSIHTMYVPYKSAKEIKTKQTSQVQNLRSIGTISPDVILCRSEVSLT
SEVKRKISLCFNVENAVNAIDLENSIYEMPLLARENIVDFLLSKFGFSKPSLDLT
DMONLVTTLCDGNRHLVRIGLVCKYLEHKDAYKSVFESLSHASTPANCSEIIVPISPE
SEALTNQLSODCGLIPGGFGTRSWEGKIAAARYCRDNNTPCFGICLGMALVVEYAR
YALUSPLANSLEMDNPTDPVCMOGODTMIRGTMRLGAYPCQITPKSLAEAYKT
DLVOERHRRHYEVNPAYVLDLQKRLRIVGVCPOGDLCEIIEIPNHRMMLGVGFHPEF
LSKLAAPHLFVKFLSAALDYSLEKGRE"
6451..6897
/gene="TC0456"
CDS
6451..6897
/gene="TC0456"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39309.1"
/db_xref="GI:7190501"
/translation="MNSMOKKEAFLGVYDKKRIGLAFASAPLLITLPIGSIINTCSSL
ALTAQALITIKERAVTVVFGNPLPMQKSYASSVQSEIOELAAALVRDMSLEVLWD
ERLSSAQAEMLKNDGCLSRKQKRNSSLSAATLILSSFLDSRLKY"
7013..8536
/gene="TC0457"
CDS
7013..8536
/gene="TC0457"
/notes="glucose-6-phosphate 1-dehydrogenase; identified by
match to PFAM protein family HMM PF00479"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="AAF73556.1"
/db_xref="GI:8163222"
/translation="MBEIEKMGPTLPACPPVMVITGATGDLTARKLPALYNLTKEG
RLSENVCVGFARRPKSHEQFREMQAIONFSHSSEIDIRVWESLEHRLIFYHQANES
EAGYSLSRFSLESVDQYKTKGNRLFLSTPPDYTFQEIIRNLRHQLFTHEOGAQOP
WSRLIIEKPFQVLDLQAEQLQCCIDANIEESYRIDHYILGKRVQNLITIRANTLIF
ESCNSQYIDHWQISYSEISIGISGNFFKSGMLDMQVHLMQLCLLTMPPEPSEF
SSAIEKKERIKILKILPIREEDYIRGOYGVQVGVSVSGVEEENVDNPSVTVV
AKLIFDNPBGVAPFYLOAKGRLPKRTTDIAVFKKSSYDLFENSENCPLCENDLLR
IIRIODEGVALQFNCKPCTKNLVRPKMDFRYSVNTVTPAYERLLCDCLIGLR
TFTSNEEVLASWELSPLEQSKIRIPFPNTIAGSLRQGADELLSRDGRWRSY"
8539..9329
/gene="TC0458"
CDS
8539..9329
/gene="TC0458"
/notes="glucosamine-6-phosphate isomerase, putative;
identified by match to PFAM protein family HMM PF01182"
/codon_start=1
```



```
/transl_table=11
/product="glucosamine-6-phosphate isomerase, putative"
/protein_id="AAE73557.1"
/db_xref="GI:8163223"
/translation="MATLISLNDANRLIAESQEDFLQIACYDWISTANKAIQKRGAF
YVALSGGKTPLOIFQETIVKRAAISDCSKIFVFWGDERASEDEAGSNYLKAMDILKW
LRTPDQIFRMDPANPKGDELYENLIREHVPDPIEDMVMLGVCDEGHTLSLFPCTAAL
EEKDRLVVNEVPQLQTRRMWTLTFPIVRQARHLVAYIQGTAKODLCHKLHLPLGRDTF
PIERVGTPLNPLQWVLSDDCCRAADLADIPAECKLEMF"
gene complement(9326..10198)
CDS /gene="TC0459"
complement(9326..10198)
```

```
Query Match 13.5%; Score 86.2; DB 1; Length 16207;
Best Local Similarity 55.9%; Pred. No. 2.2e-13;
Matches 185; Conservative 0; Mismatches 143; Indels 3; Gaps 1;

QY 7 aaaggatttttagctctcttgaggaccagagagagcagagaccagtcgttttagag 66
Db 10831 AATAGAGTGTATTGTAGTAGAGCGGAGAGGTCTGGGAAACCCCAATTTACACAG 10772

QY 67 gctctgtaccacatttagaggaagaggagtagaggtgttgacgacccgtgaacctggc 126
Db 10771 GCACCTTTCTAAACGTTTAAATGGAGAGAGGGAAAGAGGTTGTTTAAACAAGGAGCCAGGA 10712

QY 127 ggaagtcttgattgggggaagattcgcgggaagtgtatttggatccaaagtcatactcagatg 186
Db 10711 GGGAGTGCTCTTGGAGAGCAGCTTCGAGATCTTGTATTGGATGTAACACA---GGAGATA 10655

QY 187 gatgctaaacagagctactctctctatatattgccaagtgcagagacagcatttggggaaaaa 246
Db 10654 TCTTCCTACGCGGAGTTATTACTGTTCTGCGACGCGCTCAGCACATTCAAGAAAAA 10595

QY 247 gttcttcagcccttgaagctggcaagtgtggtcatcatgatcgatcgttttatcgatagttct 306
Db 10594 ATATTGCTCTCTTGAATCTGGAAAAACAGTTATATGTGACCGCTTTTCATGATCTTACT 10535

QY 307 gttgcctatcagggaatttggcgtggccttag 337
Db 10534 ATTGTATACCAAGGAATTGCTGGGGGCTAG 10504
```

Search completed: February 15, 2002, 01:16:16
Job time: 10954 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	639	100.0	639	4	US-09-259-109-1	Sequence 1, Appli
2	68.2	10.7	633	4	US-09-277-565-13	Sequence 13, Appl
3	36	5.6	7218	1	US-08-232-463-14	Sequence 14, Appl
4	32.6	5.1	2380	1	US-08-597-231-3	Sequence 3, Appli
5	31.6	4.9	143602	1	US-08-597-236-1	Sequence 1, Appli
6	31.6	4.9	14602	1	US-08-746-682A-1	Sequence 1, Appli
7	30.8	4.8	83025	1	US-08-444-734A-1	Sequence 1, Appli
8	30.6	4.8	14602	1	US-08-597-236-1	Sequence 1, Appli
9	30.6	4.8	14602	1	US-08-746-682A-1	Sequence 1, Appli
10	30.4	4.8	311	1	US-08-594-031-71	Sequence 1, Appli
11	30	4.7	1282	3	US-08-961-083-171	Sequence 71, Appl
12	30	4.7	2316	1	US-08-246-403A-7	Sequence 171, App
13	30	4.7	2316	1	US-08-246-403A-10	Sequence 7, Appli
14	29.6	4.6	2133	4	US-09-187-124-1	Sequence 10, Appl
15	29.6	4.6	5036	4	US-09-177-349-2	Sequence 1, Appli
16	29.4	4.6	2517	4	US-08-846-234-4	Sequence 2, Appli
17	29	4.5	2422	4	US-08-952-365-7	Sequence 4, Appli
18	29	4.5	3937	3	US-08-586-165-8	Sequence 7, Appli
19	28.4	4.4	289	4	US-09-007-005-17	Sequence 8, Appli
20	28.4	4.4	289	4	US-09-244-796-17	Sequence 17, Appl
21	28.2	4.4	502	4	US-09-085-199B-39	Sequence 17, Appl
22	28.2	4.4	1419	1	US-07-960-981-4	Sequence 39, Appl
23	28.2	4.4	1419	5	PCT-US93-09634-4	Sequence 4, Appli
24	28.2	4.4	8082	1	US-08-306-691B-41	Sequence 4, Appli
25	28.2	4.4	8082	1	US-08-187-785-1	Sequence 41, Appl
26	28.2	4.4	8082	5	PCT-US93-06251-28	Sequence 1, Appli
27	28.2	4.4	4403765	4	US-09-103-840A-2	Sequence 28, Appl
28	28.2	4.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli


```

; NAME/KEY: CDS
; LOCATION: 4051..4731
; OTHER INFORMATION: /product= "epsE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4898..5854
; OTHER INFORMATION: /product= "epsF"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6425..7540
; OTHER INFORMATION: /product= "epsG"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7736..8212
; OTHER INFORMATION: /product= "epsH"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8221..9192
; OTHER INFORMATION: /product= "epsI"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9285..10364
; OTHER INFORMATION: /product= "epsJ"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10392..11339
; OTHER INFORMATION: /product= "epsK"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11302..12222
; OTHER INFORMATION: /product= "CDS (eps L) covering CDS
; OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 12233..13651
; OTHER INFORMATION: /product= "epsM"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13732..14305
; OTHER INFORMATION: /function= "CDS on the
; OTHER INFORMATION: complementary strand"
; OTHER INFORMATION: /product= "orfz"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 230..252
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 274..302
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 340..345
;
; US-08-597-236-1

Query Match      4.9%; Score 31.6; DB 1; Length 14602;
Best Local Similarity 52.2%; Pred. No. 2.7;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 497 ataaaaagtcgcaaggctaccttctctcttgataaagagggaatcgcaattgtca 556
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 779 ATAGACACATATCGAGATCTTGATGTCAGCTCTCAAAAAGATATAAAAGATTGATGTTA 838

QY 557 agattgatgctagctccctttggagcaagttgtggaaactaccaaggtctgtgttg 616
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 839 AAGTTGATGATGTTGCCCTCATATCAAGAAGCTTATGATATCTCAAGCTCTGGCAATCTA 898

QY 617 acggaatgggcttg 630
   | | | | | | | | | |
Db 899 AAGCTATGGTCTTG 912

RESULT 6
US-08-746-682A-1

; Sequence 1, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746.682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 352..1803
; OTHER INFORMATION: /product= "epsA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..2535
; OTHER INFORMATION: /product= "epsB"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2547..3239
; OTHER INFORMATION: /product= "epsC"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3249..3995
; OTHER INFORMATION: /product= "epsD"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4051..4731
; OTHER INFORMATION: /product= "epsE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4898..5854
; OTHER INFORMATION: /product= "epsF"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6425..7540
; OTHER INFORMATION: /product= "epsG"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7736..8212
; OTHER INFORMATION: /product= "epsH"
; FEATURE:
; NAME/KEY: CDS

```

LOCATION: 8221..9192
OTHER INFORMATION: /product= "epsi"
FEATURE:
NAME/KEY: CDS
LOCATION: 9285..10364
OTHER INFORMATION: /product= "epsj"
FEATURE:
NAME/KEY: CDS
LOCATION: 10392..11339
OTHER INFORMATION: /product= "epsk"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 11302..12222
OTHER INFORMATION: /product= "CDS (eps L) covering CDS
OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
FEATURE:
NAME/KEY: CDS
LOCATION: 12233..13651
OTHER INFORMATION: /product= "epsM"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 13732..14305
OTHER INFORMATION: /function= "CDS on the
OTHER INFORMATION: complementary strand"
OTHER INFORMATION: /product= "orfz"
FEATURE:
NAME/KEY: terminator
LOCATION: 230..252
FEATURE:
NAME/KEY: promoter
LOCATION: 274..302
FEATURE:
NAME/KEY: RBS
LOCATION: 340..345
US-08-746-682A-1

Query Match 4.9%; Score 31.6; DB 1; Length 14602;
Best Local Similarity 52.2%; Pred. No. 2.7;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 497 ataaaaagtgcgaagctacccttctctctgataaaaggggaacgcattgtca 556
|||||
Db 779 ATAAACAATAATCGAGATCTTGATGCTCAAGCTCTCAAAAAGATAAAAAGTTGATGTTA 838
QY 557 agattgatctagctccctttggagcaagttgtgaaactaccagcgtctctctttg 616
|||||
Db 839 AAGTTCATGATGTGCTTCATATCAAGAAGCTTATGATATCTCAAGTCTGGCAAAATCTA 898
QY 617 acggaatgggcttg 630
| | | | |
Db 899 AAGCTATGGTCTTG 912
RESULT 7
US-08-444-734A-1
; Sequence 1, Application US/08444734A
; Patent No. 5610282
; GENERAL INFORMATION:
; APPLICANT: Sibley, David R.
; APPLICANT: Monsma, Frederick J.
; APPLICANT: Mahan, Lawrence C.
; APPLICANT: McVittie, Loris D.
; TITLE OF INVENTION: cDNA encoding the rat D1 dopamine
; TITLE OF INVENTION: receptor linked to adenylyl cyclase activation and
; TITLE OF INVENTION: expression of the receptor protein in plasmid-transfected
; NUMBER OF INVENTION: cell lines
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA

COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,734A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/029,917
FILING DATE: 03-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/548,714
FILING DATE: 06-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH065.001FW1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3025 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: PB73D1
FEATURE:
NAME/KEY: CDS
LOCATION: 293..1756
US-08-444-734A-1
Query Match 4.8%; Score 30.8; DB 1; Length 3025;
Best Local Similarity 53.3%; Pred. No. 2;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 442 gctgcgaatagtcacgcaggttaatcgcttgattgaaaggttgactgcataaa 501
| | | | |
Db 2069 GGTGCTAAACTCTCCGAGCAGTTTTCAGATTGGAAAGGTAGGTGATGCTTTGTTAAT 2128
QY 502 aaagttcgtcaaggctacaccttctctctgataaaagagggaatcgactgtcaagatt 561
| | | | |
Db 2129 TAACCTTCCTCAATAATAATTGAGCCTTACAGCAGGAGTGGGATTCCTTTTCTCAGAAAT 2188
QY 562 ga 563
| |
Db 2189 GA 2190
RESULT 8
US-08-597-236-1/c
; Sequence 1, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGLEE, Franscesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York


```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 352..1803
; OTHER INFORMATION: /product= "epsA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..2535
; OTHER INFORMATION: /product= "epsB"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2547..3239
; OTHER INFORMATION: /product= "epsC"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3249..3995
; OTHER INFORMATION: /product= "epsD"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4051..4731
; OTHER INFORMATION: /product= "epsE"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4898..5854
; OTHER INFORMATION: /product= "epsF"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6425..7540
; OTHER INFORMATION: /product= "epsG"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7736..8212
; OTHER INFORMATION: /product= "epsH"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8221..9192
; OTHER INFORMATION: /product= "epsI"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9285..10364
; OTHER INFORMATION: /product= "epsJ"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10392..11339
; OTHER INFORMATION: /product= "epsK"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 11302..12222
; OTHER INFORMATION: /product= "CDS (eps L) covering CDS
; OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 12233..13651
; OTHER INFORMATION: /product= "epsM"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 13732..14305
; OTHER INFORMATION: /function= "CDS on the
; OTHER INFORMATION: complementary strand"
; OTHER INFORMATION: /product= "orfz"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 230..252
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 274..302
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 340..345
; US-08-746-682A-1

```

```

Query Match 4.8%; Score 30.6; DB 1; Length 14602;
Best Local Similarity 65.2%; Pred. No. 6;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 493 ttgcataaaaaagttcgtaaggctacacctttctctcttcttgataaaagggaatcgacatt 552
||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
Db 10542 TTAATAAAAAAATCTTGAATCTACTCTTTTATCTATATGAAGAAGAAATCAAAGT 10483

QY 553 gtcaagatt 561
||| |||

Db 10482 ATTGAGAAT 10474

RESULT 10
US-08-594-031-71/c
; Sequence 71, Application US/08594031
; Patent No. 5783182
; GENERAL INFORMATION:
; APPLICANT: THOMPSON, Timothy C.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/594,031
; FILING DATE: 30-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,838
; FILING DATE: 16-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Remenick, James
; REGISTRATION NUMBER: 36,902
; REFERENCE/DOCKET NUMBER: 0A146-0110
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-594-031-71

Query Match 4.8%; Score 30.4; DB 1; Length 311;
Best Local Similarity 51.5%; Pred. No. 0.69;
Matches 70; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 294 tatcgatagttctgttcctcatcaggagttgctggtgcttagattgaagccattga 353
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 276 TATCGATGTGGGTTCTCCTAGCAGGCTCTGGATAGACGACAGACTAGTCTGCATCAT 217

QY 354 ctggctcaatcagtttgcgcagatgggtcacaacccgatttgacactattttgacat 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 216 ATAGTGCAGACCACTTATCTTGTAGGATGTCCAGGCAGATGTTGCCCTGGGTGTCACGT 157
Qy 414 cgagtggaagaagg 429
| ||||| || |||||
Db 156 TGGGTGTGTAGCAGG 141

RESULT 11
US-08-961-083-171
; Sequence 171, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 171:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-171

Query Match 4.7%; Score 30; DB 3; Length 1282;
Best Local Similarity 48.8%; Pred. No. 2.2;
Matches 81; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 404 atttgacatcgagtggaagaaggctgctgctgattgtgcgaatagtgacgcgag 463
| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1095 ATTTGAACCTATATTGAAGATTTCATGATGCTAAAGGCAATAAAATCGACCGCCTC 1154
| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Qy 464 ttaatcgcttggttggaagggttggaacttgcaataaaagttcgtcaaggctaccttt 523
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1155 CAATCAATGGAAGCACTATTGACTATTAAAGTCCACAACTGTTCAATCAGGAGACATGG 1214
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 524 ctctctggaagaagggaatcgcatgttgcgaattgctag 569
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1215 TTCGAGCTCTTAAGAGGGCTTATCAATCTTTATAAGGAGATGG 1260
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-246-403A-7
; Sequence 7, Application US/08246403A
; Patent No. 5583040
; GENERAL INFORMATION:
; APPLICANT: Kaji, Akira

```

```

; TITLE OF INVENTION: Mutation of RepA Protein
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,403A
; FILING DATE: 20-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30330
; REFERENCE/DOCKET NUMBER: 2020-102P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2316 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 182..571
; OTHER INFORMATION: /product= "Repa from Pl, see Fig.
; OTHER INFORMATION: 9"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 891..1748
; OTHER INFORMATION: /product= "IncC from Pl, see Fig.
; OTHER INFORMATION: 9"
; US-08-246-403A-7

Query Match 4.7%; Score 30; DB 1; Length 2316;
Best Local Similarity 50.7%; Pred. No. 3.2;
Matches 72; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 453 tgaccgcgaggttaactgttggttggaagggttggaacttgcaataaaagttcgtca 512
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1607 TGTGGCGACCAACCGTATCATTAAGAAGCGATAGACAGCTTAACACATCGGCTA 1666
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 513 aggtacattctctctgtgataaagagggaatcgcatgttgcgaattgatgctagct 572
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1667 TCTCGACTGTCTATTGAGAAGAAAGGCCGGGAAAGTTTGTATCTCCATTCTCGAA 1736
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 573 cccttgaggacaggttgga 594
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1727 TCCAAAGCTGAAACTCCCCGAA 1748
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-246-403A-10
; Sequence 10, Application US/08246403A
; Patent No. 5583040
; GENERAL INFORMATION:
; APPLICANT: Kaji, Akira
; TITLE OF INVENTION: Mutation of RepA Protein
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747

```

Search completed: February 15, 2002, 01:49:59
Job time: 4772 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 22:03:21 ; Search time 1181.61 Seconds
(without alignments)
5811.183 Million cell updates/sec

Title: US-09-749-972-1

Perfect score: 639

Sequence: 1 atgtcaaaagatttttagt.....gaatgggcttgccaaatga 639

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_est1: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	51.4	8.0	251	10	AU063540	AU063540	AU063540
C 2	42.6	6.7	1050	13	CNS012BP	AL101359	Drosophil
C 3	38.6	6.0	682	10	AL509002	AL509002	AL509002
C 4	38.6	6.0	1092	13	CNS04YG5	AL312926	Tetraodon
C 5	38.4	6.0	1101	13	CNS00LT2	AL078714	Drosophil
C 6	36.8	5.8	646	13	A2814364	A2814364	2M081J23
C 7	36.2	5.7	707	11	BF158171	BF158171	fl29e03.y
C 8	36	5.6	535	11	BG691744	BG691744	341040 BA
C 9	35.6	5.6	562	10	AU154589	AU154589	AU154589
C 10	35.6	5.6	758	13	AQ858832	AQ858832	nbe0003B
C 11	35.4	5.5	299	13	AQ882061	AQ882061	HS_5337_A
C 12	35.4	5.5	464	11	BG566008	BG566008	602583190

13	35.4	5.5	480	10	A1021635	A1021635	u13f11.r
14	35	5.5	1101	13	CNS0106X	AL098595	Drosophil
C 15	34.8	5.4	516	10	AW306038	AW306038	fj79b05.y
C 16	34.8	5.4	521	10	BE201309	BE201309	fk89q01.y
C 17	34.8	5.4	839	13	CNS0314D	AL248998	Tetraodon
C 18	34.6	5.4	608	10	AU039978	AU039978	AU039978
C 19	34.6	5.4	682	10	AU039910	AU039910	AU039910
C 20	34.6	5.4	689	11	C90165	C90165	C90165
C 21	34.4	5.4	364	11	BG979728	BG979728	CM4-CN009
C 22	34.4	5.4	401	10	AA549022	AA549022	vk77e03.s
C 23	34.4	5.4	613	11	BF231990	BF231990	df11e10.x
C 24	34.4	5.4	619	10	AA546725	AA546725	vk66e07.s
C 25	34.2	5.4	878	13	CNS05EJI	AL333783	Tetraodon
C 26	34.2	5.4	1008	11	BF259733	BF259733	HVSMF002
C 27	34.2	5.4	1201	13	CNS015W6	AL105984	Drosophil
C 28	34	5.3	562	11	B1425539	B1425539	saf32f06.
C 29	33.8	5.3	474	11	BG578759	BG578759	de35e03.y
C 30	33.8	5.3	515	10	A1727581	A1727581	BNLGH1847
C 31	33.8	5.3	595	10	AW640503	AW640503	bl95H04.y
C 32	33.8	5.3	599	11	BG162656	BG162656	de35e03.y
C 33	33.8	5.3	618	13	A2874122	A2874122	2M0188F10
C 34	33.8	5.3	646	10	A1727515	A1727515	BNLGH1831
C 35	33.8	5.3	1101	13	CNS000D1	AL065414	Drosophil
C 36	33.6	5.3	470	10	AA050890	AA050890	mj22h12.r
C 37	33.6	5.3	775	13	A2127961	A2127961	OSJNB006
C 38	33.4	5.2	580	10	A1881565	A1881565	606070G11
C 39	33.4	5.2	859	13	CNS004YY	AL055406	Drosophil
C 40	33.2	5.2	454	13	A2049503	A2049503	GSSBRU010
C 41	33.2	5.2	632	10	AU221260	AU221260	AU221260
C 42	33.2	5.2	727	11	C89525	C89525	C89525
C 43	33.2	5.2	924	11	BF535569	BF535569	602051860
C 44	33.2	5.2	1072	13	CNS0390U	AL234183	Tetraodon
C 45	33	5.2	357	10	AW687837	AW687837	NF014A06R

ALIGNMENTS

RESULT 1
LOCUS AU063540 251 bp mRNA EST 30-APR-2001
DEFINITION AU063540 Rice callus Oryza sativa cDNA clone C62637, mRNA sequence.
ACCESSION AU063540
VERSION AU063540.2 GI:13895842
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
AUTHORS 1 (bases 1 to 251)
TITLE Sasaki, T. and Minobe, Y.
JOURNAL Rice cDNA from callus
COMMENT Unpublished (1994)
On May 20, 1999 this sequence version replaced gi:4879241.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.
C62637_12Z.
Location/Qualifiers
1. .251
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C62637"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using

36 a	79 c	82 g	54 t

Query Match	8.0%	Score 51.4;	DB 10;	Length 251;
Best Local Similarity	51.1%	Pred. No. 0.00027;		
Matches 121; Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0;

[illegible]

Qy 167 atccaagtcatactcagatggatgctaaacagagctacttctctatatgccagtcqca 226

525 CACCCTGGCTCCGGGCGATCTCCCTGCTGAGGCTTTGCTGTACAAACGCCGACAAGG 127

[illegible]

Qy 287 atcgttttatcgcgatagttctgtgtgcctatcagggaatttggtcgtggcttagatatg 343

.....

CNS012BP/c

NOTES AND CORRECTIONS

ACCESSION
VERSION

SOURCE

REFERENCE

JOURNAL

COMMENT

SECRET

0
0
4
5
0
2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

NOTES

Best Local

295 atco
Qy

1

```

/clone="029015"
/clone_lib="B"
/note="Genoscope sequence ID : COAB029AH08B3-end : SP6"
BASE COUNT      239 a    236 c    266 g    298 t    53 others
ORIGIN

Query Match          6.0%; Score 38.6; DB 13; Length 1092;
Best Local Similarity 60.2%; Pred. No. 2;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY  4 tcacaaagatttttagtctcttgaggaccagagggagcaggcaagaccagtgttta 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  907 TGAANAGCCTTTGTGTCGTTCCCGCGCANAAAGTGGGAGCAGAACCCGCAGTGTTCG 848
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  64 gaggtctgtaccaatatttagaggaagaaaaggagtga 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  847 AAACGCTGCTGGATAGTTCATAGAAAAAAGCNTAAA 810
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
CONSOLUT2          CNS00LT2       1101 bp      DNA           GSS        14-JUN-1999
LOCUS              Drosophila melanogaster genome survey sequence TET3 end of BAC;
DEFINITION         BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
                    fly), genomic survey sequence.
ACCESSION          AL078714
VERSION            GI:5102004
KEYWORDS            GSS; fruit fly.
SOURCE             Drosophila melanogaster
ORGANISM            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE           1 (bases 1 to 1101)
AUTHORS            Direct Submission
TITLE              Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage ;
JOURNAL            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                   - Web : www.genoscope.cns.fr)
COMMENT            Determination of this BAC-end sequence was carried out as part of a
                   collaboration with The Berkeley Drosophila Genome Project (BDGP).
                   The BDGP is constructing a physical map of the Drosophila
                   melanogaster genome using these BACs. For further information
                   please see http://www.fruitfly.org The BDGP Drosophila
                   melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                   Aaron Mammoser in Pieter de Jong's laboratory in the Department of
                   Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                   NY. The library is named RPCI-98 and was constituted by partial
                   EcoRI digestion of Drosophila DNA provided by the BDGP from the
                   isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
                   pl and EST libraries. A more detailed description of the library
                   and how to order individual BAC clones, the entire library, or
                   filters for hybridization from the BACPAC Resource Center can be
                   found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source             1..1101
                   /organism="Drosophila melanogaster"
                   /db_xref="taxon:7227"
                   /clone_lib="RPCI-98"
                   /clone="BACR48P19"
                   /note="end : TET3"
BASE COUNT         469 a     6 c     69 g     151 t     406 others
ORIGIN

Query Match          6.0%; Score 38.4; DB 13; Length 1101;
Best Local Similarity 19.0%; Pred. No. 2.3;
Matches 75; Conservative 134; Mismatches 185; Indels 0; Gaps 0;

QY  233 atttgtggaagaagtcttcacgcccttgagctgcgaagtggatcgatcatcgatcgtt 292
   | | : | : | | | | | | : | : | : | : | : | : | : | : | : | : |
Db  566 AKTAAKTAAGAAWTTTTTAAATTTTNTKTWTAKTKWAKWAWMDTATTWTTWTTT 625

```

QY	293	tta	c	a	g	a	t	a	g	t	t	c	t	c	t	a	c	a	g	g	a	t	t	c	g	t	c	g	t	a	t	t	c	a	c	a	t	t	g	352
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	626	W	T	T	T	T	T	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
QY	353	a	c	t	c	g	c	a	a	t	c	a	g	t	t	c	a	c	a	g	a	t	t	c	a	c	a	g	a	t	t	c	a	c	a	t	t	g	412	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
Db	686	W	T	T	D	T	W	A	A	A	W	T	T	K	K	K	K	K	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
QY	413	t	c	a	g	a	t	t	c	a	a	g	t	c	g	t	a	t	t	c	g	a	a	t	a	g	a	c	c	a	g	a	t	t	a	a	t	472		
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
Db	746	K	G	G	K	K	G	K	G	K	K	K	K	K	K	K	K	G	G	K	K	K	K	A	G	D	D	D	A	K	D	K	T	K	K	K	A			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
QY	473	t	g	a	t	t	c	a	a	g	t	t	c	a	a	a	a	a	a	a	a	a	a	t	c	g	t	c	a	a	g	t	c	a	c	c	t	532		
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
Db	806	K	K	K	G	K	K	K	A	A	K	K	A	A	D	P	T	K	K	T	W	D	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
QY	533	a	t	a	a	a	g	a	a	a	t	c	a	g	a	t	t	c	a	a	a	t	t	c	a	a	g	a	t	t	c	a	a	a	t	t	592			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
Db	866	K	K	K	G	G	K	D	A	A	A	K	K	K	G	T	K	K	K	K	K	G	K	G	K	G	K	G	K	G	K	G	K	G	K	G	K			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
QY	593	a	a	a	c	t	a	a	a	g	c	t	c	t	t	c	t	t	t	t	t	a	c	a	a	t	t	t	t	t	a	c	a	a	t	t	626			
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
Db	926	G	G	G	K	K	G	K	G	K	G	K	G	K	K	T	K	T																						

RESULT	6
AZ814364/c	
LOCUS	AZ814364 646 bp DNA
DEFINITION	2M0081J23R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUGC2M0081J23 R, DNA sequence.
	20-FEB-2001 GSS

VERSION AZ814364.1 GI:12984188

KEYWORDS
SOURCE
house mouse.

REFERENCE 1 (bases 1 to 646)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@gene

Insert Length: 10

Plate: 0081 Row: 1
Seq primer: CACAC

Seq primer: CACAC

High quality sequ

FEATURES

```
source 1. .646
```

/organism

/strain=

```

/ab_xref
/ab_xref="

```

clone/
clone/
clone=

```

/CRONE_I
/sex="Ma

```

/lab_host

/note="V

musculus

Laboratory

(http://www.bud...

was injured

1

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pW42 (gi|4732114|gb|AF19207.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor-mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	170 a	161 c	132 g	183 t
ORIGIN				

```
Query Match          5.8%; Score 36.8; DB 13; Length 646;
Best Local Similarity 44.5%; Pred. NO. 5.8;
```

Qy 1 atgtcaaaaggatttttagtctctcttgaggagccagagggagcagggcaagaccagtggtt 60
|||||
Db 591 AAGTCAGTATGTGTGTGCAGCAGAAATGGGGGGAAATTCACAGCAGTGGATTTCTCTGGT 532
|||||

Qy 61 tttagaggctgctgctaccacatttttagagggaataaaggagtagagggtgttgacaccccdtgaa 120

Db 531 CCTGAGCCTGTGGTGGCAGACGTGTTGAGGTGGGACAAAGCCAGGCATGATTCCTTAA 472

Qy 121 cctggcggactcttgattggggagaagaattcgggaagtgtatttggatccaagtcaatact 180
|||
Db 471 ACTTATATGCAATTTACATGCCTCCAAATGCCTCCTAGAGGCTGTCTGATCAGCCTCTGT 412
|||

Qy 181 cagatggatgctaaacagagctactctctatatattgccagtcgcagacacagcatttgggtg 240
Db 411 CTTAAGGTAGGGTGACACATACTTCGAATTTTGGATACCAGGACACAGCTTTACTTT 352

Qy 241 gaaaaatcttcagcccttgaagtgcacatcatgatcgttttatcgat 300
 ||||| ||||| ||||| ||||| |||||
Db 351 TAAAAGAAATCGGGGCTTTGGACTGGGATATTTGTAGTGTGTAGATGC***GCCCT 292

Qy	301	agttctgttgccctatcaggatttggtc	328
Db	291	AGTACCGTTGAAGCCCTGGGATTGATC	264

RESULT 7
BF158171/c

LOCUS	BF158171	707 bp	mrna	EST	23-FEB-2001
DEFINITION	f129e03.y1 Sucano kawakami zebrafish DRA Danio rerio cDNA clone 3815141 5' similar to TR:Q9Y3W8 Q9Y3W8 HYPOTHETICAL 71.4 KD PROTEIN ;, mRNA sequence.				

ACCESSION BF158171
 VERSION BF158171.1 GI:11053375
 KEYWORDS EST.
 SOURCE zebrafish.

ORGANISM	Danio rerio
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Rasbora; Danio.
REFERENCE	1 (bases 1 to 707)

AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Willson, R.

TITLE	WASNU Zebrafish EST Project 1998
JOURNAL	
COMMENT	Unpublished (1998) Contact: Stephen L. Johnson Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@rri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES
source
1..562
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cloned_lib="NT2RP4001564"
/cclone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/note="Vector: pMEI8SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 167 a 135 c 88 g 166 t 6 others
ORIGIN

Query Match 5.6%; Score 35.6; DB 10; Length 562;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 104; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY	404	atttgacatcgaggtggaagaaggctgcgtcgtattgttcgcaatatgtagccgcgag	463
Db	237	ATATTGATGAAGTGGCGAAGAATAATTGGAGATTCTGGCTAACGACGGCTGGTCAAGATG	178
QY	464	ttaactcgttggatttggaagggttggaacttgcataaaaagtctcgaaggtcaccttt	523
Db	177	GAGATGATGAAGATTGGGAAGAGATGATGCTGAANAGACTGCTCTGGAAGGCTATTTCN	118
QY	524	ctctctcggataaaggaaatcgcatctgcaagattgatgctagctctccccttggagc	583
Db	117	CAATCATGTGATGAAGAGATAACCCTGTTGATGAGTATCAGATATTTAAAGCTATCTTTC	58
QY	584	aagttgtggaactaccaggctgtctgtttgacggaatg	624
Db	57	NAACTATTCAAATCGTAATCCCTGTGGTATCAGGCACITG	17

RESULT 10
AQ58832/c
LOCUS
DEFINITION
nbeo003B09r CUGI RNA BAC Library (EcoRI) Oryza sativa genomic clone nbeb0003B09r, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.
1 (bases 1 to 758)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAAACAGCTATCACCATG
Class: BAC ends
High quality sequence start: 31
High quality sequence stop: 503.
Location/Qualifiers
1..758
/organism="Oryza sativa"

FEATURES
source

100

```

/clone_lib="Soares_mammary_gland_NbMWG"
/sex="male"
/tissue_type="mammary_gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: p7T3D-Pac (Pharmacia
RI) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTCACCAATCTCAAGTCGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      111 a      141 c      130 g      98 t
ORIGIN

Query Match      5.5%; Score 35.4; DB 10; Length 480;
Best Local Similarity 47.2%; Pred. No. 13;
Matches 108; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 97 gtagagggttgacgacccgtgaacctgcgcgagctcttgattggggagaagattcgggaa 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 GTGAGCTCGGGTGCAITTCAGGACCAGCTGGAGGACTGCTCCATCATCATCTTCACCAA 129
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 157 gtgatttgatccaagtcatactcagatgatgactaaaacagatcactctctatatt 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 TACCTGCAAGACCTGCGAGATCCTTTGTATGATGCTTCGAAAGTTCAACTTCCCCACAGT 189
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 217 gccagtcgcagacgacatttggtggaaaaagtcttccagcccttgaagctggcgaattg 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GGCACTGCATCCATGATGAAGCAGAAAGAGCGCTTTGGCCCTTGGCAGATTCAGTCC 249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 277 gtcatcagatcgctttatcagatagctctgttgccctatcaggagatttg 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 AGCATCTATCGGATTCATTATTGCCACTGACGTGCGCTCAAGGGGCGCTGG 298
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
CNS0106X
LOCUS
DEFINITION
BACNO3K20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL098595
VERSION        AL098595.1 GI:5610206
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS       Genoscope.
TITLE         Direct Submission
JOURNAL       Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BU 191 9106 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAcl1.

FEATURES
Source         1. ..1101
                /organism="Drosophila melanogaster"
                /plasmid="pBelOBAcl1"

```

```

Seq primer: T3 ET from Amersham
High quality sequence stop: 433.
Location/Qualifiers
1. .516
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish gridded kidney"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOLR"
/note="Organ: kidney; Vector: pBK-CMV; Site.1: ECORI;
Site.2: XhoI; Oligo dt cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."
BASE COUNT      138 a      124 c      126 g      127 t      1 others
ORIGIN

```

```

Query Match      5.4%; Score 34.8; DB 10; Length 516;
Best Local Similarity 50.6%; Pred. No. 20;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 27  tgaggaccagagggcaggcaagaccagtgttttagaggctctgtaccatttaga 86
      || || || || || || || || || || || || || || || || || ||
Db 203  TGCAGGCCACAGAACTCCAGAGCTGAGTCCGGTCTGCTCTCTCTCTCTCTGT 144

QY 87  ggaataaggagtagaggtgttgacgacccgtgaacctggcggagctcttgattggggagaa 146
      || || || || || || || || || || || || || || || || || ||
Db 143  TGAGCAAAGAGCTGATGCTACTGATGTAACTGCTCCAGTTGGACTGATGCTGAGGAGGG 84

QY 147  gattcgggaagtgtatttgatccaagtcaatcactcagatggatgct 192
      || || || || || || || || || || || || || || || || || ||
Db 83  CCAGCAGGTATGGCCTGTATGAGGAGGCCACACTCACACCAGTCT 38

```

Search completed: February 15, 2002, 00:50:14
Job time: 10013 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2002, 23:40:32 ; Search time 102.36 Seconds
(without alignments)
5352.001 Million cell updates/sec

Title: US-09-749-972-1
Perfect score: 639
Sequence: 1 atgtcaaaagatttttagt.....gaatgggttgcccaatga 639

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_1101.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	639	100.0	639	21	AAA64398
2	635.8	99.5	639	21	AA291859
3	635.8	99.5	28171	19	AAV52155
4	634.2	99.2	1632	19	AAZ96394
5	131	20.5	840	21	AAZ51229
6	125.8	19.7	5855	20	AAZ13042
7	119.4	18.7	4029	18	AAV74612
8	107.2	16.8	49617	22	AAZ28541
9	89.6	14.0	1038602	20	AAZ01425
c 10	86.6	13.6	8102	20	AAI13181
11	81.6	12.8	349980	21	AAZ21608

12	68.2	10.7	633	20	AA237114	Nucleotide sequenc
13	56.6	8.9	58073	18	AA58840	Mycoplasma genital
c 14	42.4	6.6	3332	22	AAH54919	S. epidermidis gen
c 15	39.4	6.2	936	22	AAF58252	Oligonucleotide D1
c 16	39.4	6.2	936	22	AAF58254	Oligonucleotide D1
c 17	39.4	6.2	936	22	AAF58257	Oligonucleotide D1
c 18	39.4	6.2	936	22	AAF58259	Oligonucleotide D2
c 19	39.4	6.2	936	22	AAF58262	Oligonucleotide D2
c 20	39.4	6.2	936	22	AAF58255	Oligonucleotide D1
c 21	39	6.1	936	22	AAF58252	Oligonucleotide D1
22	39	6.1	936	22	AAF58254	Oligonucleotide D1
23	39	6.1	936	22	AAF58257	Oligonucleotide D1
24	39	6.1	936	22	AAF58259	Oligonucleotide D2
25	39	6.1	936	22	AAF58262	Oligonucleotide D2
26	39	6.1	938	22	AAF58255	Oligonucleotide D1
27	35.4	5.5	805	22	AAH05496	Human cDNA clone (
28	35.4	5.5	1793	22	AAH15903	Human cDNA sequenc
29	33.8	5.3	194	22	AAI21374	Probe #11307 for g
30	33.8	5.3	194	22	AAI46655	Probe #15341 used
31	33.8	5.3	194	22	AAI07069	Probe #7060 used t
32	33.8	5.3	847	22	AAH08614	Human cDNA clone (
33	33.8	5.3	2291	22	AAH16769	Human cDNA sequenc
34	32.8	5.1	32768	20	AAI12954	Enterococcus faeca
35	32.8	5.1	62909	22	AAF28545	Genomic fragment #
c 36	32.6	5.1	465	22	AAF94354	Fusarium venenatum
c 37	32.6	5.1	583	21	AAF09313	Haemophilus influe
c 38	32.6	5.1	759	20	AAI16593	Fusarium venenatum
c 39	32.6	5.1	2380	19	AAV66834	Human gene express
c 40	32.6	5.1	2478	18	AAI67287	Zea mays solubie s
41	32.2	5.0	1605	21	AAI42127	Soluble starch syn
c 42	32.2	5.0	3263	22	AAH16211	Arabidopsis thalia
43	32	5.0	3333	21	AAI59415	Human cDNA sequenc
44	31.6	4.9	14602	18	AAI73236	Human secreted pro
c 45	31.6	4.9	15449	21	AAH1526	Streptococcus ther
						N. meningitidis pa

ALIGNMENTS

RESULT 1
AAA64398
ID AAA64398 standard; DNA; 639 BP.
AC AAA64398;
DT 20-DEC-2000 (first entry)
XX DNA encoding a thymidylate kinase polypeptide.
DE Thymidylate kinase; tdk polypeptide; bacterial infection; ss.
KW Thymidylate kinase; tdk polypeptide; bacterial infection; ss.
XX Streptococcus pneumoniae.
OS
FH Key Location/Qualifiers
CDS FT 1..639
FT /*tag= a
FT /product= "thymidylate kinase"
XX
PN WO2000050602-A1.
XX
XX 31-AUG-2000.
XX
XX 18-FEB-2000; 2000WO-US04238.
XX
XX 26-FEB-1999; 99US-0259109.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Burnham MKR, Zalacain M, Biswas S, Chalker AF, Ingraham KA;
PI Traini CM, Warren PV;
XX WPI; 2000-533181/48.
DR P-PSDB; AAB08637.

XX Nucleic acids encoding thymidylate kinase family polypeptides derived
PT from *Streptococcus pneumoniae*, useful for screening for antibacterial
PT agents -
XX
XX
PS Claim 2; Page 2-3; 4lpp; English.
XX
CC The present sequence encodes a polypeptide of the thymidylate kinase
CC family (tdk polypeptides), and is derived from *Streptococcus pneumoniae*.
CC Tdk polynucleotides may be used to recombinantly produce the tdk
CC polypeptides either *in vivo* (e.g. as part of a genetic vaccination
CC procedure) or *in vitro* (e.g. as part of a fermentation culture).
CC The nucleic acids and proteins may be used to diagnose diseases
CC in which the tdk polypeptides are expressed, such as infection by
CC *Streptococcus pneumoniae*.
XX
XX Sequence 639 BP; 165 A; 114 C; 179 G; 181 T; 0 other;
SQ

Query Match 100.0%; Score 639; DB 21; Length 639;
Best Local Similarity 100.0%; Pred. No. 3.7e-192;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgtcaaaaggatttttagtctctcttgagggaaccagagggagcaggcaagaccagtggt	60
Db	1	atgtcaaaaggatttttagtctctcttgagggaaccagagggagcaggcaagaccagtggt	60
Qy	61	ttagaggctcgtaccaaatttagagaaaaaggagtagaggtgttgacgaccggtgaa	120
Db	61	ttagaggctcgtaccaaatttagagaaaaaggagtagaggtgttgacgaccggtgaa	120
Qy	121	ccctggcgagtccttgattgggggagaagattcgggaagtattttggtaccaagtcatact	180
Db	121	ccctggcgagtccttgattgggggagaagattcgggaagtattttggtaccaagtcatact	180
Qy	181	cagatggatgctaaaaacagagctacttctctatatattgccagtcgacagacatttggtg	240
Db	181	cagatggatgctaaaaacagagctacttctctatatattgccagtcgacagacatttggtg	240
Qy	241	gaaaaagtcttcagccctgaagctggcaagctgtgtcatcatggatcggtttttatcgat	300
Db	241	gaaaaagtcttcagccctgaagctggcaagctgtgtcatcatggatcggtttttatcgat	300
Qy	301	agttctgttgcctatcagggaatttggctgtggtctagatatattgaagccaattgactggctc	360
Db	301	agttctgttgcctatcagggaatttggctgtggtctagatatattgaagccaattgactggctc	360
Qy	361	aatcagtttgcacagatgggtccaaaccgattttgacactctattttgacatcagggtg	420
Db	361	aatcagtttgcacagatgggtccaaaccgattttgacactctattttgacatcagggtg	420
Qy	421	gaagaaggctcgctcgattgctgcgaatagtgaccgcgaggttaactcgtttgatttg	480
Db	421	gaagaaggctcgctcgattgctgcgaatagtgaccgcgaggttaactcgtttgatttg	480
Qy	481	gaagggttgacttgcataaaaaagttcgtcaaggctaccttctctcttgatgataaagag	540
Db	481	gaagggttgacttgcataaaaaagttcgtcaaggctaccttctctcttgatgataaagag	540
Qy	541	ggaaatcgatattgcaagattgatctctccctttggagcaagttgtgaaactacc	600
Db	541	ggaaatcgatattgcaagattgatctctccctttggagcaagttgtgaaactacc	600
Qy	601	aaggctgtcttgttgacggaatgggcttggccaatga	639
Db	601	aaggctgtcttgttgacggaatgggcttggccaatga	639

RESULT 2

AAZ91859
ID AAZ91859 standard; DNA; 639 BP.

AA
AC AA291859;

1

XX	02-JUN-2000	(first entry)
XX	Streptococcus pneumoniae	DNA sequence ID119.
XX	Streptococcus pneumoniae	infection; immunogen; antigen; diagnosis; AIDS;
XX	bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;	
XX	kidney disease; diabetes; immunosuppressive disorder; otitis media;	
XX	pneumococcal septicaemia; sinusitis; meningitis; therapy; ss.	
XX	Streptococcus pneumoniae.	
XX	WO200006738-A2.	
XX	10-FEB-2000.	
XX	27-JUL-1999;	99WO-GB02452.
XX	27-JUL-1998;	98GB-0016336.
XX	19-MAR-1999;	99US-0125329.
XX	(MICR-) MICROBIAL TECHNIQS LTD.	
XX	Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;	
XX	WPI; 2000-195301/17.	
XX	P-PSDB; AAY81763.	
XX	Streptococcal proteins and polynucleotides useful for diagnosis,	
XX	treatment and prophylaxis of bacterial infections	-
XX	Claim 2; Page 61; 76pp; English.	
XX	This sequence encodes a Streptococcus pneumoniae protein of the	
XX	invention. The proteins (or their homologues, derivatives and/or	
XX	fragments)are useful as immunogens or antigens. Immunogenic or antigenic	
XX	compositions comprising the proteins are useful as vaccines and also in	
XX	diagnostic assays. The sequences are useful for the detection or	
XX	diagnosis of S. pneumoniae infection, by contacting a sample to be tested	
XX	with them. Agents capable of antagonising, inhibiting or interfering with	
XX	the function or expression of the protein or polypeptide are useful in	
XX	medical compositions in the treatment or prophylaxis of S. pneumoniae	
XX	infection. As the sequences can be used to treat S. pneumoniae infection,	
XX	they can be used to treat bacterial pneumonia, which has high rates in	
XX	young children, the elderly, and in patients with predisposing conditions	
XX	such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,	
XX	or with immunosuppressive disorders, especially AIDS. They can also be	
XX	used to treat pneumococcal septicaemia, otitis media, sinusitis, and	
XX	meningitis.	
XX	Sequence 639 BP; 165 A; 115 C; 177 G; 182 T; 0 other;	
XX		

Query Match	99.5%	Score 635.8;	DB 21;	Length 639;
Best Local Similarity	99.7%	Pred. No. 3.8e-191;		
Matches 637: Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

[illegible]


```
CC The present sequence is also useful for detecting Staphylococcal
XX infections in biological samples.
SQ Sequence 840 BP; 332 A; 93 C; 170 G; 245 T; 0 other;

Query Match      20.5%; Score 131; DB 21; Length 840;
Best Local Similarity 55.2%; Pred. No. 2.3e-31;
Matches 323; Conservative 0; Mismatches 250; Indels 12; Gaps 3;

QY 15 tttagtctcttgaggaccagagagagcagacagaccagtggttttagagcctgct 74
   ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 ttttataacttttgaggcccaagagctgctggaaaacaaactgtaattaagaagttta 169

QY 75 accaattttagaggaagaaaggagtagagggttgacgacccgtgaacctggcggagcttt 134
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 170 ccatagatita--gtaaaagattatgatgtcattatgactagagaccaggtgtgttcc 226

QY 135 gattggggagaagattcgggaagtgattttggatccaagtcatactacagatgactaa 194
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 tactgtgagaagaatacgtaaaaattgtattagaaggcaatgat-----atggacattag 280

QY 195 aacagagctactctctatattccagtcgacagcagcatttggtgaaaaagtctctcc 254
   ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 281 aactgaagcaatgttattgtctcatctcagagagacatctgtattaaaggctacacc 340

QY 255 agccctgaagctggcgaagtgtgctcatctatggatcgttttatcgatagttctgtgcta 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 agctttaaaagaaggttaaggttgtgtgtgatcgctatatcgatagttcattagctta 400

QY 315 tcagggatttgcgtggcttagatattgaagccattgactggctcaatcagtttgcgac 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 tcaaggttatgctagagggttgcgttgagaagtaagcattaaacgaatttgcgaat 460

QY 375 agatgggctcaaacccgatttgacactctattttgacatcgagtggaagaaaggctggc 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 aaatgattatataccagacttgcagatttattttaaattgttagtgcgaagtagtcgcga 520

QY 435 tcgtattgtcgaatagtagccgcgaggtttaatcgttttgatttggaaaggttggactt 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 acgtattatt--aaaaattcaagagatcaaaatagattagatcaagaagatttaaaagt 577

QY 495 gcataaaaaagtcgtcaaggctacatttctctcttcttggtataagagggaaatcgcatgt 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 tcacgaaaagaagtaatgaaggttaccgaagaatacattcataatgaatcacaacggttcaa 637

QY 555 caagatgatgctagctctcccttggagcaagttgtggaaactac 599
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 aagcgttaatgcagatcaaacctcttgaaaatgttgttgaagacac 682

RESULT 6
AAI13042
ID AAI13042 standard; DNA; 5855 BP.
XX
AC AAI13042;
XX
DT
XX
19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:105.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
WO9850555-A2.
XX
PN
XX
12-NOV-1998.
PD
XX
04-MAY-1998; 98WO-US08985.
XX
14-NOV-1997; 97US-0066009.
XX
```

```
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 676-679; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAI12938 to AAI13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 5855 BP; 1856 A; 1061 C; 1221 G; 1713 T; 4 other;

Query Match      19.7%; Score 125.8; DB 20; Length 5855;
Best Local Similarity 60.3%; Pred. No. 2.5e-29;
Matches 223; Conservative 0; Mismatches 145; Indels 2; Gaps 1;

QY 8 aaggatttttagtctctcttgaggaccagagggagcagcagcagtggttttagagg 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5486 agggattttttatcacatttgaaaggtccagatggtgctggtataaacacagtggtttgaagg 5545

QY 68 c--tcgtctaccatttttagaggaagaaaggagtagagggttgacgacccogtgaacctgg 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5546 aagtttctgaanattagctaaagagtcacaaacgaaaaattgtcactacaagagaaccagg 5605

QY 126 cggagcttcttgaggagagattcgggaagtgatttggatcccaagtcatactcagat 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5606 cggcataccgattgccgaaagattcgaacagttatttttagatccaaganatgcagaat 5665

QY 186 ggatgctaaaaacagagctactctctatattgccagtcgcagacagcatttggtgaaaaa 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5666 ggatgagcgaacagaagaagcttactttacgcagctgcgtcgtcgaacatttagtggaaaa 5725

QY 246 agttctccagcccttgaagctggcaagttggtcctatcatgatgctgtttttatcgatgttc 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5726 aatatgccagcttggaaagcgggtcatttagtcctatgctgtatgctgtttgtgacagctc 5785

QY 306 tgttgctatcaggagatttggctgcttagatattgaagccattgactggctcaatca 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5786 attagcctaacaggcgtgacgtngaatggggatgggcccccatgcctcaattaatgc 5845

QY 366 gtttgcgaca 375
   | | | | | | | | | |
Db 5846 ggttgcggca 5855

RESULT 7
AAV74612/c
ID AAV74612 standard; DNA; 4029 BP.
XX
AC AAV74612;
XX
```

16-MAR-1999 (first entry)
 Staphylococcus aureus contig SEQ ID #301.
 Computer readable medium; vaccine; S.aureus infection; Immunodetection;
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 skin infection; surgical wound infection; scalded skin syndrome;
 toxic shock syndrome; ds.
 Staphylococcus aureus.
 Key Location/Qualifiers
 misc_feature 181..240
 /tag= a
 /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 misc_feature 1981..2040
 /tag= b
 /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 misc_feature 3781..3840
 /tag= c
 /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 EP786519-A2.
 30-JUL-1997.
 07-JAN-1997; 97EP-0100117.
 05-JAN-1996; 96US-0009861.
 (HUMA-) HUMAN GENOME SCI INC.
 Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 Rosen CA;
 WPI; 1997-374922/35.
 Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1; Page 1130-1133; 3271pp; English.
 This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S.aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S.aureus infection. The
 polypeptides can also be used in a kit for the immunodetection of
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S.aureus DNA sequences contained on the
 computer readable medium.
 Sequence 4029 BP; 1110 A; 793 C; 552 G; 1388 T; 186 other;

Query Match 18.7%; Score 119.4; DB 18; Length 4029;
 Best Local Similarity 55.0%; Pred. No. 2.2e-27;
 Matches 322; Conservative 0; Mismatches 251; Indels 12; Gaps 4;
 QY 15 tttagctctcttgaggagcagagggagcagagcagctgttttagaggctctgt 74
 DB 1257 TTTTATAACTTTTGGAGGCCAGAGAGGCTCTGGAAGAAACACTGTAATTAATGAAGTTA 1198
 QY 75 accaattttagagaaaaagagtagaggtgttgacgagcccgtagaacctgcggaggtctt 134
 DB 1197 CCAATAGATTA---GTAAAAGATTTATGATGTCATTATGACTAGAGAACCCAGGTGGTTC 1141
 QY 135 gattggggagagattcgggaagtgtatttgatccaagtcatcatcagatgagtgctaa 194
 DB 1140 TACTGGTGAAGNAATACGTAAATGTATTAGAAAGG-----CAATGATATGGACATTTAG 1086
 QY 195 aacgagagtaactctctatattgcagtcgcagacagcatttgggtggaaaaagttctcc 254
 DB 1085 AACTGAAGCAATGTTATTTGCTGCATCTAGAAGAGAACATCTTGTTATTAAGGTCTATACC 1026
 QY 255 agccctgaagctggcgaagtggctcatcatgagatgtttatcagatagttctgtgctcta 314
 DB 1025 AGCTTTAAAGAAAGTAAGGTGTGTTGCTGATCGCTATATCGATAGTTCATTAGCTTA 966
 QY 315 tcagggtattggtcgtggttagattattgaagccattgactggtcactcagtttgcgac 374
 DB 965 TCAAGGTATCTAGAGGGATGGCGTTGAGAGAGTAAAGACANTTAACGAAATTTGCAAT 906
 QY 375 agatgggctcaaacccgatttgacactctatttgacatcgaggtgggaagaggtctgac 434
 DB 905 AATGGATTATATCCAGACTTGACGATTTATTTAAATGTTAGTGTGAAGTAGGTGCGGA 846
 QY 435 tcgtatgctgcgaatagtagccgcgaggttaactgctgttggatttggaggggttggactt 494
 DB 845 ACGTATTATT---AAAAATTCAGAGATCAAAATAGATTAGATCAAGAAAGATTTAAAGTT 789
 QY 495 gcataaaaaagttcgtcaaggctaccttctctctggtgataaaggaggaatcgcattgt 554
 DB 788 -CAGGAAAAAGTAATTTGAAGGTTACCAAGAAATCATTCATTAATGAATCACAACGGTTCAA 730
 QY 555 caagattgatgctagtcctcccttggagcgaagtgttggaactac 599
 DB 729 AAGCGTTAATGCAGATCAACCTCTTGAAAATGTTGTTGAAGACAC 685
 RESULT 8
 AAF28541
 ID AAF28541 standard; DNA; 49617 BP.
 XX AC AAF28541;
 XX DT 04-APR-2001 (first entry)
 XX DE Genomic fragment #28.
 XX KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 XX KW bronchopulmonary; endocarditis; meningitis; ss.
 XX OS Moraxella catarrhalis.
 XX PN WO200078968-A2.
 XX PD 28-DEC-2000.
 XX PF 16-JUN-2000; 2000WO-US16649.
 XX PR 18-JUN-1999; 99US-0140121.
 XX PA (INCY-) INCYTE GENOMICS INC.
 XX PI Lagace RE, Patterson C, Berg KL;
 XX XX

DR	WPI; 2001-041427/05.
XX	
PT	Genomic library for identifying diagnostic and therapeutic
PT	compositions, and for identifying virulence factors, regulatory
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic
PT	acids -
XX	
XX	
PS	Claim 1; Page 235-247; 545pp; English.
XX	
CC	The present invention relates to a Moraxella catarrhalis genomic library
CC	comprising of a combination of 41 nucleic acid molecules (see
CC	AAF28514-AAF28554). The library has a number of uses described in the
CC	specification e.g. is useful for identifying diagnostic and therapeutic
CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC	aerobic, gram-negative diplococcus, normally found among the bacterial
CC	flora of human upper airways. M. catarrhalis is known to cause acute,
CC	localised infections such as otitis media, sinusitis and bronchopulmonary
CC	infection and life-threatening, systemic diseases including endocarditis
CC	and meningitis.
XX	
XX	Sequence 49617 BP; 14572 A; 9853 C; 11026 G; 14166 T; 0 other;

	Query Match	16.8%	Score 107.2;	DB 22;	Length 49617;
	Best Local Similarity	50.8%;	Pred. No. 5.4e-23;		
	Matches 309;	Conservative 0;	Mismatches 293;	Indels 6;	Gaps 2;
QY	8	aaggatttttagtctcttgaggacagagagagagcaggaacagcagctgttttaaggg	67		
Db	41430	aaggaaatttaccgttttgagggtactgagcggtgggtaaaaccacgcgaattgaag	41489		
QY	68	ctctgcaccaattttagagaaaaagagtagagggtgttgacgacccgtgaacctggcg	127		
Db	41490	cactcatctcgttttggtgatcgtggcacaagtcgtacgcacagagagccaggtg	41549		
QY	128	gagcttgattggggagaagattcgggaagtgttttggatcccaagtcatactcagatcg	187		
Db	41550	gtagtgaagctggccgagcagttgcgtcaactatttttaaatccga---atgccaatatga	41606		
QY	188	atgctaaaacagagctactctcttatattccagctcgacagacagcatttgggtgaaaaag	247		
Db	41607	atgctgacacagagattttgctgatgtttgctgcacgcagtgatcatttgcacaggtga	41666		
QY	248	ttcttcagcccttgaagctggcaagttggtctcatcattggtctttttatcatagttctg	307		
Db	41667	tttacctgcaattaaatcgggttaaatgggtggtcagtgaccgctttttgatgtagtcgg	41726		
QY	308	tgtcctatcagggatttggctgttgattgatatattgaagccattgactggctcaatcagt	367		
Db	41727	tggcttatcaggggttttggacgatatcatgaaacaaagcgcaattggctgaagatttgac	41786		
QY	368	ttgcacagatgggctccaaacccgatttgaca---ctctattttgacatcgaggtggaag	424		
Db	41787	ttttgattgacagtttggcccaatttaccagatgtaacaccttggcttgatttggatg	41846		
QY	425	aagggtcgctcgctattgctgcgaatagtgaccgcgaggtttaatcgttttgatttggaa	484		
Db	41847	tgggtatgggtatgcagcgtgcgcgcaaaagaaagtgtgctgatcgtattggaagccaacg	41906		
QY	485	ggttgacttgcataaaaaagttcgtcaaggtacaccttctctcttggtataaaagggaa	544		
Db	41907	acattgcttttttttgagcgtgtgtatcaaggtttggccatcatcagcacgacgactcca	41966		
QY	545	atcgcatgtcaagattgatgtcagttctccttttggagcaagttgtggaaactaccaagg	604		
Db	41967	atcgattatcgcattgatgcgaatggcactgtggcagaagttggcggcacgcatagata	42026		
QY	605	ctgtcttg 612			
Db	42027	ccgactttg 42034			

RESULT 9

AAZ01425	
ID	AAZ01425 standard; DNA; 1038602 BP.
XX	
AC	AAZ01425;
XX	
DT	07-OCT-1999 (first entry)
XX	
DE	Complete genome sequence of Chlamydia trachomatis.
XX	
KW	Vaccine; eye disease; conventional trachoma; nonendemic trachoma; paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis; nongonococcal urethritis; epididymitis; cervicitis; salpingitis; Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX	
OS	Chlamydia trachomatis.
XX	
PN	WO9928475-A2.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-TB01939.
XX	
PR	04-NOV-1998; 98US-0107077.
PR	28-NOV-1997; 97FR-0015041.
PR	17-DEC-1997; 97FR-0016034.
XX	
PA	(GEST) GENSET.
XX	
PI	Griffais R;
XX	
DR	WPI; 1999-371125/31.
XX	
PT	Genome sequence of Chlamydia trachomatis
XX	
PS	Claim 1; Page 373-656; 1755pp; English.
XX	
CC	The present sequence represents the complete genome of Chlamydia trachomatis. Open reading frames (ORFs) of the genome encode polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perihepatitis, Bartholinitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.
XX	
SO	Sequence 1038602 BP: 304265 A: 214645 C: 214259 G: 305001 T: 432 other:

Query Match	14.0%	Score 89.6;	DB 20;	Length 1038602;
Best Local Similarity	57.7%;	Pred. No. 8.4e-17;		
Matches 180;	Conservative 0;	Mismatches 129;	Indels 3;	Gaps 1;
QY 26	ttgaggaccagagggagcagacgagctgttttagagcgtctgtctaccaattttgag	85		
Db 264145	tagaaggcggaaggtgtctgggaaactcagttattatacagcgtttcttaagcgttttaa	264204		
QY 86	agaaaaagagtagaggtgttgacgacccctgaaacctggcggagcttggattgggggaga	145		
Db 264205	tagaagaaggcagagaaattgtctactacgagagcagcggggatgctctctttggagaca	264264		
QY 146	agattcgggaagtgtatttggatccaaagtactactcagatggatggtgctaaaaacagactac	205		
Db 264265	gtgttcaggcctctcttggaatccagaacaaa--aaatatccatatgcagaattgt	264321		
QY 206	ttctctatttgcagtcgcagacagcatttgggtggaaaaagtttcttcagccctttgaag	265		
Db 264322	tgttgttttagcagctctgcacacacattcaggaaaaagatcatacctgctcttaagt	264381		
QY 266	ctggcaagtgtgctcatcatgagtcggtttttatccgatagtctctgttgcctatcaggagatttg	325		

[illegible]

Db	8041	TAGACGTTGACTCAGATACCTGGCTTACGCGCGGATTCAAGAAATCGAACACAGSAAATTG	7982
QY	467	atcgtttgattggaagggttgagacttgacataaaaaagttcgtaaggctacattttctc	526
Db	7981	ATCGCTTAGATTCAGAAAGGATTAGAATTTCATCAACGKGTTGCGTCATGAATATTGAAAT	7922
QY	527	tctgataaagagggaatcgcattgtcaagattgatgtcagtcctcccttggagcaag	586
Db	7921	TACGAGAGAAATCCACACGCTATCAAAAAGATTGTGCACGGATGAGTTTAGAACTTG	7862
QY	587	ttdtgaaactacaaaggctgtcttgttga	617
Db	7861	TCGTAGAAGCAACGTTATCAGCGGATTTATGA	7831
RESULT	11		
AAF21608			
ID	AAF21608	standard; DNA; 349980 BP.	
XX	AAAF21608;		
AC	AAAF21608;		
XX			
DT	13-MAR-2001	(first entry)	
XX			
DE	Neisseria meningitidis B	nucleotide sequence SEQ ID NO:109.	
XX			
KW	Neisseria meningitidis; Neisseria gonorrhoeae;	immunogenic; vaccine;	
KW	diagnosis; antigen; detection; infection;	gene therapy; antibacterial;	
XX	ds.		
XX			
OS	Neisseria meningitidis.		
OS			
PN	W0200066791-A1.		
XX			
PD	09-NOV-2000.		
XX			
PF	08-MAR-2000; 2000WO-US05928.		
XX			
PR	30-APR-1999; 99US-0132068.		
PR	08-OCT-1999; 99WO-US23573.		
PR	28-FEB-2000; 2000GB-0004695.		
XX			
PA	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX			
PI	Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;		
PI	Galeotti C, Mora M, Rattli G, Scarselli M, Scarlato V, Rappuoli R;		
PI	Frazer CM, Grandi G;		
XX			
DR	WPI; 2000-647603/62.		
XX			
PT	Neisseria meningitidis B	full length genome sequence and open reading	
PT	frames are used to detect, treat and prevent Neisserial infections -		
XX			
PS	Claim 7; Appendix A; 692pp; English.		
XX			
CC	The present invention describes the full length genome of		
CC	Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607		
CC	to AAF21613 represent fragments of the NMB genomic sequence, as the		
CC	sequence was too long to go in a record on its own it was split into 8		
CC	sequences which overlap each other at the beginning and end of each		
CC	sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at		
CC	the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at		
CC	the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the		
CC	Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to		
CC	AAF21606 represent PCR primers which are used in the exemplification of		
CC	the present invention. The NMB genome and fragments from it have		
CC	antibacterial activity, and can be used in vaccines and gene therapy.		
CC	Neisseria nucleic acids, proteins and/or antibodies which binds to the		
CC	proteins can be used in compositions for treating or preventing infection		
CC	due to Neisserial bacteria or as a diagnostic reagent for detecting the		
CC	presence of Neisserial bacteria or of antibodies raised to Neisserial		
CC	bacteria. Computers, computer memory, computer storage medium or computer		

KW M. genitalium; DNAA; DNA gyrase; origin of replication;
 XX megabase shotgun sequencing method; open reading frame; ORF; ss.
 OS Mycoplasma genitalium.
 FH Key Location/Qualifiers
 FT 8552..9184
 FT /tag= a
 FT /label= MG006
 FT /note= "Previously identified as MORF-20076, the
 FT encoded protein shows 27.59 percentage
 FT identity to thymidylate kinase (CDC8)
 FT from Saccharomyces cerevisiae"
 FT 11252..12040
 FT /tag= b
 FT /label= MG009
 FT /note= "Previously identified as MORF-20078, the
 FT encoded protein shows 35.43 percentage
 FT identity to the Bacillus subtilis hypothetical
 FT protein covered in accession number
 FT GB:D26185_102"
 FT 12069..12725
 FT /tag= c
 FT /label= MG010
 FT /note= "Previously identified as MORF-20079, the
 FT encoded protein shows 25.73 percentage
 FT identity to DNA primase (dnaE) from
 FT Clostridium acetobutylicum"
 FT complement (13570..14247)
 FT /tag= d
 FT /label= MG012
 FT /note= "Previously identified as MORF-20080, the
 FT encoded protein shows 31.50 percentage
 FT identity to the ribosomal protein S6
 FT modification protein (rimK) from Escherichia
 FT coli"
 FT complement (14396..15217)
 FT /tag= e
 FT /label= MG013
 FT /note= "Previously identified as MORF-19823, MORF-20080
 FT and MORF-20081, the encoded protein shows 33.04
 FT percentage identity to 5,10-methylene-tetra-
 FT hydrofolate dehydrogenase (fold) from E. coli"
 FT 17474..19243
 FT /tag= f
 FT /label= MG015
 FT /note= "Previously identified as MORF-20084, the
 FT encoded protein shows 32.23 percentage
 FT identity to transport ATP-binding protein
 FT (msbA) from E. coli"
 FT 26478..27344
 FT /tag= g
 FT /label= MG023
 FT /note= "Previously identified as MORF-20092, the
 FT encoded protein shows 45.96 percentage
 FT identity to fructose-bisphosphate aldolase
 FT (tsr) from B. subtilis"
 FT 27345..28448
 FT /tag= h
 FT /label= MG024
 FT /note= "Previously identified as MORF-19826 and
 FT MORF-20093, the encoded protein shows 46.84
 FT percentage identity to GTP-binding protein
 FT from E. coli"
 FT 36987..38978
 FT /tag= i
 FT /label= MG032
 FT /note= "Previously identified as MORF-20099, the
 FT encoded protein shows 26.82 percentage
 FT identity to ATP-dependent nuclease (addA)
 FT from B. subtilis"
 FT 39242..39904
 FT /tag= j

FT /label= MG033
 FT /note= "Previously identified as MORF-20100, the
 FT encoded protein shows 35.90 percentage
 FT identity to glycerol uptake facilitator
 FT (glpF) from B. subtilis"
 FT complement (39873..40514)
 FT /tag= k
 FT /label= MG034
 FT /note= "Previously identified as MORF-20101, the
 FT encoded protein shows 48.13 percentage
 FT identity to thymidylate kinase (tdk)
 FT from B. subtilis"
 FT 40543..41787
 FT /tag= l
 FT /label= MG035
 FT /note= "Previously identified as MORF-20102, the
 FT encoded protein shows 30.71 percentage
 FT identity to histidyl-tRNA synthetase (hiss)
 FT from Mycobacterium leprae"
 FT complement (44751..46277)
 FT /tag= m
 FT /label= MG038
 FT /note= "Previously identified as MORF-20105, the
 FT encoded protein shows 46.83 percentage
 FT identity to glycerol kinase (glpK)
 FT from E. coli"
 FT complement (46268..47422)
 FT /tag= n
 FT /label= MG039
 FT /note= "Previously identified as MORF-19831 and
 FT MORF-20106, the encoded protein shows 43.20
 FT percentage identity to glycerol-3-phosphate
 FT dehydrogenase (GUT2) from S. cerevisiae"
 FT 49377..49643
 FT /tag= o
 FT /label= MG041
 FT /note= "The encoded protein shows 48.86 percentage
 FT identity to phosphohistidinoprotein-hexose
 FT phosphotransferase (ptsH) from Mycoplasma
 FT capricolum"
 FT 50060..51520
 FT /tag= p
 FT /label= MG042
 FT /note= "Previously identified as MORF-19832 and
 FT MORF-20108, the encoded protein shows 41.92
 FT percentage identity to spermidine/
 FT putrescine transport ATP-binding protein
 FT (potA) from E. coli"
 FT 51525..52382
 FT /tag= q
 FT /label= MG043
 FT /note= "Previously identified as MORF-20110, the
 FT encoded protein shows 26.51 percentage
 FT identity to spermidine/putrescine transport
 FT system permease protein (potB) from E. coli"
 FT 52366..53220
 FT /tag= r
 FT /label= MG044
 FT /note= "Previously identified as MORF-20111, the
 FT encoded protein shows 29.45 percentage
 FT identity to spermidine/putrescine transport
 FT system permease protein C (potC) from E. coli"
 FT 54658..55605
 FT /tag= s
 FT /label= MG046
 FT /note= "Previously identified as MORF-20112, the
 FT encoded protein shows 36.60 percentage
 FT identity to sialoglycoprotease (gcp)
 FT from Pasteurella haemolytica"
 FT complement (56970..58310)
 FT /tag= t
 FT /label= MG048
 FT /note= "Previously identified as MORF-19834,

QY	1	atgcaaaagattttttagctctctctgaggaccagaggagcaggcagacacagtgtt	60
Db	8552	atgaataaagggtgtttgtttatgaaggagttgatggagcgggcaaacgtcttta	8611
QY	61	ttagaggctctgtaccaaatttagaggaaaaagagtagagggtgtgac-----gac	113
Db	8612	atgtagaggt-tttaaaaactttatccaaactaagtttttgaaactatcaactta	8670
QY	114	ccgtgaacctgcggagcttctgattggggagagaagattcgggaagtgatttggatccaag	173
Db	8671	tagagaacctggtagtcttctgtagtcgaaaaaattcgtcaactctctttaaatgaaa-	8729
QY	174	tcaactcagatggatgctataaacagagcctactctctctatttgcgcagtcgcagacagca	233
Db	8730	-----caatggaacctcaactgaagcttatttgttgcgcagctagaactgaa	8781
QY	234	tttggcgaaaagtctctccagcccttgaagctggcgaagttggtcatcagggatcgttt	293
Db	8782	tatcagtagcgttaattcaaacccgcaattgaaaaagacaacactagttatttcagatagatt	8841
QY	294	tatcgtatgtctctgttcgtatcaggagatttggctggtggtatattgaagccattga	353

Db	8842	tgtttctctagtttgcataccaaggattaaagcaaaaaataagcattgatacagtaaa	8901
Qy	354	ctggctcaatcagtttgcacagatgggctcaaacccgatttgacactctatttgacat	413
Db	8902	acagattaatcatcatcggttaagaataatgatgcaaaccttatttttggattg	8961
Qy	414	cgaggttgaagaaggctggctgctgatttctgctgcgaaatagtgaccgcgagggttaatcgttt	473
Db	8962	caatttaagaagcattacaaggatgcaaaagcgttgtaataatactcttgatga	9021
Qy	474	ggatttggaaaggttgacttcataaaaaaatttcgtcaaggctacettctcttcgga	533
Db	9022	atttataagaagaagaatgatttgatacagttcgtctcttattatttaagcttagtga	9081
Qy	534	taaaagggaaatcgcttgcataagattgatgcta	568
Db	9082	taaaaaaactgttcttgattaatgggataata	9116
RESULT	14		
ID	AAH54919/C		
XX	AAH54919 standard; DNA; 3332 BP.		
XX	AAH54919;		
AC			
03-SEP-2001	(first entry)		
XX			
DT			
DE	S. epidermidis genomic polynucleotide sequence SEQ ID NO:4283.		
XX			
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;		
KW	vaccination; endocarditis; ds.		
XX			
OS	Staphylococcus epidermidis.		
XX			
PN	WO200134809-A2.		
XX			
PD	17-MAY-2001.		
XX			
PF	09-NOV-2000; 2000WO-US30782.		
XX			
PR	09-NOV-1999; 99US-0164258.		
XX			
PA	(GLAX) GLAXO GROUP LTD.		
XX			
PI	Kimmerly WJ;		
XX			
DR	WPI; 2001-316495/33.		
XX			
PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,		
PT	useful for vaccinating against infections, e.g. endocarditis -		
XX			
PS	Claim 8; Page 2014-2015; 2188pp; English.		
XX			
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides		
CC	(II), given in AA881454 to AA883120, from Staphylococcus epidermidis.		
CC	(I) and (II) can have antibacterial activity and therefore can be used		
CC	in vaccination. The nucleic acids (I) may be used to produce the		
CC	S. epidermidis polypeptides (II) via the production of vectors		
CC	containing them which are used to produce hosts cells which express the		
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be		
CC	used to vaccinate subjects and to raise antibodies against the bacteri		
CC	The polypeptides may also be used to assay for other inhibitors of the		
CC	activity and therefore identify compounds that may be used for the		
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to		
CC	AAH53090 represent specifically claimed S. epidermidis genomic DNA		
CC	polynucleotide sequences from the present invention. AAH55091 to		
CC	AAH55098 represent oligonucleotide sequences and primers which are		
CC	in the exemplification of the present invention.		
CC	N.B. The present invention specifically claims all the polynucleotide		
CC	sequences given in the sequence listing of the present specification,		
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even		
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4477		
CC	no sequences are present for SEQ ID NO:4455 to 4464.		

```

XX      SQ      Sequence 3332 BP; 982 A; 685 C; 487 G; 1178 T; 0 other;

Query Match          6.6%;   Score 42.4;   DB 22;   Length 3332;
Best Local Similarity 53.0%;   Pred. No. 0.0047;
Matches 141; Conservative 0; Mismatches 116; Indels 9; Gaps
                                |
QY      16      ttgagtctctcttgagggaccagaggagcagcgcaagaccagtgcttttagagagctctgcta 75
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      259      TTTTATACGTTTGAAGTCCAGAGGAGGAGCAGGAAACACCTCAGTAATTAAAAAGTTAAGC 200

QY      76      ccaatttttagaggaanaagagtagagggtgttgacgaccgtgaacctggcggagtcctg 135
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      199      GAAAGATTA---GCTAAAGAATATGATATTGTTTACAACACAGCTGAACCTGGTGGAGTATTA 143

QY      136      attggggagaaagtctcggaagtgatttggatcccaagtcatactcagatggtatgctctaaa 195
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      142      ACTAGTGAAGAAATACGTCGATTGTGTACGTGTATGGAGATAGTATAGACAT-----TCGA 89

QY      196      acagagctactctctctatattccagtcgcagacagcatttggtgaaaaaagtctctcca 255
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      88      ACTGAAGCTATGCTGTTTCGCTGCATCTAGACGTGAACACCTTGTGTGAAAGATTATCCCA 29

QY      256      gcoccttgaagctggcgaagtttggtcat 281
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      28      TCTTTACAGGCGAGGTAAAAATTGTCCT 3
                                ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT	15
AAF58252/c	
ID	AAF58252 standard; DNA; 936 BP.
XX	
XX	AAF58252;
XX	AC
XX	AC
DT	24-APR-2001 (first entry)
XX	
XX	Oligonucleotide D1835.
DE	
XX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
KW	
XX	Synthetic.
OS	
XX	WO200107665-A2.
XX	
PN	01-FEB-2001.
PD	
XX	
XX	26-JUL-2000; 2000WO-US20476.
PF	
XX	
XX	26-JUL-1999; 99US-0145695.
PR	
PR	17-MAR-2000; 2000US-0190259.
XX	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	
XX	
XX	Umek RM;
PI	
PI	
DR	WPI; 2001-159728/16.
XX	
XX	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
PT	
XX	
PS	Example 6; Page 127; 159pp; English.
XX	
XX	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
XX	Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:16:17 ; Search time 37.91 Seconds
(without alignments)
414.232 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKGLVSLGPEGAGTSTV.....LEQVVETTKAVLFDGMGLAK 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	100.0	212	21	AAB08637
2	1053	100.0	212	21	AAY81763
3	1047	99.4	212	19	AAY86073
4	528	50.1	212	20	AAY28785
5	475.5	45.2	214	22	AAB72200
6	475	45.1	205	21	AAY70146
7	354	33.6	206	20	AAY37013
8	329.5	31.3	209	20	AAY34868
9	306	29.1	210	20	AAY28787
10	305.5	29.0	213	20	AAY28786
11	305.5	29.0	213	22	AAB72201

12	300	28.5	210	20	AAY28792	Thymidylate kinase
13	277	26.3	210	20	AAY28791	Thymidylate kinase
14	200	19.0	188	20	AAY28789	Thymidylate kinase
15	198.5	18.9	191	19	AAW98749	H. pylori GHPO 104
16	185.5	17.6	216	20	AAY28796	Yeast thymidylate
17	185.5	17.6	216	22	AAB72202	S. cerevisiae thym
18	160.5	15.2	224	21	AAG20878	Arabidopsis thalia
19	160.5	15.2	224	21	AAG50881	Arabidopsis thalia
20	160.5	15.2	263	21	AAG20877	Arabidopsis thalia
21	160.5	15.2	263	21	AAG50880	Arabidopsis thalia
22	130	12.3	334	22	AAG84942	Shrimp white spot
23	129	12.3	159	21	AAG20879	Arabidopsis thalia
24	129	12.3	159	21	AAG50882	Arabidopsis thalia
25	123.5	11.7	227	20	AAY28790	Thymidylate kinase
26	121.5	11.5	212	20	AAY28788	Human thymidylate
27	111.5	10.6	204	20	AAY28794	Thymidylate kinase
28	111.5	10.6	227	11	AA807563	Polypeptide encode
29	110.5	10.5	205	20	AAY28795	Thymidylate kinase
30	103	9.8	571	22	AA16781	Peptide #3215 enco
31	103	9.8	571	22	AA29266	Peptide #3303 enco
32	103	9.8	571	22	AA04498	Peptide #3180 enco
33	102	9.7	54	22	AA370417	A. actinomycetemco
34	102	9.7	203	22	AG90585	C. glutamicum prote
35	102	9.7	203	22	AAB80129	Corynebacterium gl
36	101.5	9.6	210	20	AAY28793	Schizosaccharomyce
37	97.5	9.3	488	22	AA31470	Amino acid sequenc
38	97	9.2	265	20	AA03752	S. aureus Ferrichr
39	91	8.6	181	22	AG90365	C. glutamicum prote
40	91	8.6	181	22	AA80115	Corynebacterium gl
41	90.5	8.6	287	21	AA82698	Aeropyrum pernix p
42	90	8.5	448	20	AA34908	Chlamydia pneumoni
43	89.5	8.5	488	22	AA31471	Amino acid sequenc
44	89.5	8.5	1045	20	AA07482	Human chondrocyte-
45	89.5	8.5	1045	21	AAV91947	Human cytoskeleton

ALIGNMENTS

RESULT 1
AAB08637
ID AAB08637 standard; Protein; 212 AA.
XX
AC AAB08637;
XX
DT 20-DEC-2000 (first entry)
XX
DE Amino acid sequence of a thymidylate kinase polypeptide.
XX
KW Thymidylate kinase; tdk polypeptide; bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO200050602-A1.
XX
PD 31-AUG-2000.
XX
PF 18-FEB-2000; 2000MO-US04238.
XX
PR 26-FEB-1999; 99US-0259109.
XX
PR (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Burnham MKR, Zalacain M, Biswas S, Chalker AF, Ingraham KA;
PI Traini CM, Warren PV;
XX
DR WPI; 2000-533181/48.
XX
DR N-PSDB; AAA64398.
XX
PT Nucleic acids encoding thymidylate kinase family polypeptides derived
PT from Streptococcus pneumoniae, useful for screening for antibacterial
PT agents -
XX

PS Claim 1; Page 3; 4lpp; English.

XX The present sequence represents a polypeptide of the thymidylate kinase

CC family (tdk polypeptides), and is derived from Streptococcus pneumoniae.

CC tdk polynucleotides may be used to recombinantly produce the tdk

CC polypeptides either in vivo (e.g. as part of a genetic vaccination

CC procedure) or in vitro (e.g. as part of a fermentation culture).

CC The nucleic acids and proteins may be used to diagnose diseases

CC in which the tdk polypeptides are expressed, such as infection by

CC Streptococcus pneumoniae.

XX

SQ Sequence 212 AA;

Query Match 100.0%; Score 1053; DB 21; Length 212;

Best Local Similarity 100.0%; Pred. No. 1.1e-103;

Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKGLVSLGEGAGKTSVLEALLPILKEKGVEVLTTRPGGVILGKIREVILDPST 60

Db 1 mskglvslgpegagktsvleallpilleekgvevlttrepggvilgkirevildpsht 60

Qy 61 QMDAKTELLLYASRROHLVEKVLPALEAGKLVIMDRFIDSSVAYQVGRGLDTEADWL 120

Db 61 qmdakteillyasrrqlhvekvlpaleagklvimdrfidssvayqvgfgrldtealdwl 120

Qy 121 NOFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLEGLDHLKRVQGYLSLLDKE 180

Db 121 nqfatdglkpdltlyfdieveeeglariaansdrevnrlidlegldhlhkvrgyyslldke 180

Qy 181 GNRIVKIDASLPLEQVETTKAVLFDGMGLAK 212

Db 181 gnrvikidaspleqvettkavifdgmglak 212

RESULT 2

AAy81763

ID AAY81763 standard; Protein; 212 AA.

XX

AC AAY81763;

XX

DT 02-JUN-2000 (first entry)

XX

DE Streptococcus pneumoniae protein sequence ID119.

XX

KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;

KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;

KW kidney disease; diabetes; immunosuppressive disorder; otitis media;

KW pneumococcal septicaemia; sinusitis; meningitis; therapy.

XX

OS Streptococcus pneumoniae.

XX

XX WO200006738-A2.

XX

PD 10-FEB-2000.

XX

XX 27-JUL-1999; 99WO-GB02452.

XX

XX 27-JUL-1998; 98GB-0016336.

PR

PR 19-MAR-1999; 99US-0125329.

XX

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX

XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

XX

XX WPI; 2000-195301/17.

DR

DR N-PSDB; AAZ91859.

XX

XX Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections

XX

PS Claim 2; Page 61; 76pp; English.

XX

CC This sequence represents a Streptococcus pneumoniae protein of the

CC invention. The proteins (or their homologues, derivatives and/or

CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic

CC compositions comprising the proteins are useful as vaccines and also in

CC diagnostic assays. The sequences are useful for the detection or

CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested

CC with them. Agents capable of antagonising, inhibiting or interfering with

CC the function or expression of the protein or polypeptide are useful in

CC medical compositions in the treatment or prophylaxis of S. pneumoniae

CC infection. As the sequences can be used to treat S. pneumoniae infection,

CC they can be used to treat bacterial pneumonia, which has high rates in

CC young children, the elderly, and in patients with predisposing conditions

CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be

CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and

CC meningitis.

XX

SQ Sequence 212 AA;

Query Match 100.0%; Score 1053; DB 21; Length 212;

Best Local Similarity 100.0%; Pred. No. 1.1e-103;

Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKGLVSLGEGAGKTSVLEALLPILKEKGVEVLTTRPGGVILGKIREVILDPST 60

Db 1 mskglvslgpegagktsvleallpilleekgvevlttrepggvilgkirevildpsht 60

Qy 61 QMDAKTELLLYASRROHLVEKVLPALEAGKLVIMDRFIDSSVAYQVGRGLDTEADWL 120

Db 61 qmdakteillyasrrqlhvekvlpaleagklvimdrfidssvayqvgfgrldtealdwl 120

Qy 121 NOFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLEGLDHLKRVQGYLSLLDKE 180

Db 121 nqfatdglkpdltlyfdieveeeglariaansdrevnrlidlegldhlhkvrgyyslldke 180

Qy 181 GNRIVKIDASLPLEQVETTKAVLFDGMGLAK 212

Db 181 gnrvikidaspleqvettkavifdgmglak 212

RESULT 3

AAy86073

ID AAY86073 standard; Protein; 212 AA.

XX

AC AAY86073;

XX

DT 10-APR-2000 (first entry)

XX

DE S. pneumoniae derived protein #282.

XX

XX Treatment; prevention; disease; diagnosis; gene therapy; screening;

KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.

XX

OS Streptococcus pneumoniae.

XX

XX WO9806734-A1.

XX

PD 19-FEB-1998.

XX

XX 15-AUG-1997; 97WO-US14436.

PF

PR 16-AUG-1996; 96US-0024022.

XX

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

PI Stodola RK;

XX

XX WPI; 1998-159452/14.

DR

DR N-PSDB; AAZ96394.

XX

XX Streptococcus pneumoniae proteins and related DNA - useful for

PT

WPI; 1999-508654/42.

New molecule/molecule

XX
PT
New molecule/molecular complex. useful for designing inhibitors of

PT Staphylococcus aureus thymidylate kinase (Tmk), comprises a S. aureus
PT Tmk or adenosine triphosphate binding pocket -
PS Disclosure; Page 285-286; 288pp; English.
XX
CC This invention relates to a molecule or molecular complex comprising a
CC Staphylococcus aureus thymidylate kinase (Tmk) thymidine 5' monophosphate
CC (TMP) or adenosine 5' triphosphate (ATP) binding pocket. The
CC specification includes protein co-ordinate data for thymidylate kinase.
CC The x-ray crystal structure of S. aureus Tmk is useful for solving the
CC structure of other molecules/molecular complexes. It is also useful for
CC designing inhibitors of S. aureus Tmk activity. These inhibitors are
CC useful for inhibiting the growth of S. aureus or causing its death. They
CC can be used for the prevention and treatment of S. aureus Tmk mediated
CC disease and therefore exhibit antibacterial activity. Further uses
CC include the development of chemical entities that can isomerise to short
CC lived reaction intermediates in the chemical reaction of a substrate or
CC other compound that interferes with S. aureus Tmk. The reaction
CC intermediates of S. aureus Tmk can be deduced and can be used to design
CC improved analogues of known S. aureus Tmk inhibitors or to design new
CC inhibitors with high specificity and stability based on the reaction
CC intermediates. The present sequence represents a recombinant S. aureus
CC thymidylate kinase protein with a polyhistidine tag.
XX
SQ Sequence 214 AA;

Query Match 45.2%; Score 475.5; DB 22; Length 214;
Best Local Similarity 49.0%; Pred. No. 2.3e-42;
Matches 98; Conservative 40; Mismatches 57; Indels 5; Gaps 4;
Qy 1 MSKGLVLEGGEGAGKTSVLEALLPILEKGVLEVTTRPGGVLIGKIREVILDPSTHT 60
Db 1 mgsaf-itfedpegsgkttvnevyrhl-vkdyvmtrepvgvptgeeikrvlegn-- 56
Qy 61 QMDAKTELLYTASRROHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWL 120
Db 57 dmdirtamlfaasrrhlvlkvpalkvegkvvlcdryidslayqgyargigveevral 116
Qy 121 NOFATDGLKPLDLYFIEVEEGLARIAANSRDREYNRLDLEGLDHLKVRQGYLSLLDKE 180
Db 117 nefalnglypdltylnvsaevgreriikns-rdgnrlldqedlkfhekviqyqeihne 175
Qy 181 GNRIVKIDASLPLEQVETT 200
Db 176 sgrfksvndqplenvvdt 195

RESULT 6
AAY70146
ID AAY70146 standard; Protein; 205 AA.
XX AAY70146;
XX-
XX- 06-JUN-2000 (first entry)
XX
DE Staphylococcus aureus thymidylate kinase Tmk.
XX
XX Vaccine; antibacterial; prevention; attenuation; detection;
KW Staphylococcal infection; neonatal conjunctivitis; skin infection;
KW toxic shock syndrome; osteomyelitis; Tmk; thymidylate kinase.
XX
OS Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH Domain 11..14
FT Domain /label= Antigenic_epitope
FT Domain 147..149
FT Domain /label= Antigenic_epitope
XX
XX WO200012678-A2.
XX
XX 09-MAR-2000.

XX 31-AUG-1999; 99WO-US19726.
XX
XX 01-SEP-1998; 98US-0098964.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bailey CC, Choi GH;
XX
XX WPI; 2000-237864/20.
XX
XX N-PSDB; AAZ51229.
XX
XX Staphylococcus aureus polypeptide useful for preventing or attenuating
XX a Staphylococcal infection comprises one of 32 sequences of 100-1277
XX amino acids or their fragments -
XX
XX Claim 9; Page 17; 144pp; English.
XX
XX The present sequence is a thymidylate kinase Tmk
XX from Staphylococcus aureus genomic DNA library.
XX The present sequence is useful in preparation of vaccines for
XX prevention or attenuation of Staphylococcal infections (especially
XX S.aureus infections) which may cause conditions such as neonatal
XX conjunctivitis, osteomyelitis, skin infections and toxic shock syndrome.
XX The present sequence is also useful for detecting Staphylococcal
XX infections in biological samples.
XX
SQ Sequence 205 AA;
Query Match 45.1%; Score 475; DB 21; Length 205;
Best Local Similarity 49.5%; Pred. No. 2.4e-42;
Matches 96; Conservative 40; Mismatches 54; Indels 4; Gaps 3;
Qy 7 VSLEGEGAGKTSVLEALLPILEKGVLEVTTRPGGVLIGKIREVILDPSTHTOMDAKT 66
Db 5 itfepegsgkttvnevyrhl-vkdyvmtrepvgvptgeeikrvlegn--dmdirt 61
Qy 67 ELLYIASSRROHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNOFATD 126
Db 62 eamfafaasrrhlvlkvpalkvegkvvlcdryidslayqgyargigveevralnefain 121
Qy 127 GLKPLDLYFIEVEEGLARIAANSRDREYNRLDLEGLDHLKVRQGYLSLLDKEGNRIVK 186
Db 122 glypdltylnvsaevgreriikns-rdgnrlldqedlkfhekviqyqeihnesqrfs 180
Qy 187 IDASLPLEQVETT 200
Db 181 vndqplenvvdt 194
RESULT 7
AAY37013
ID AAY37013 standard; Protein; 206 AA.
XX AAY37013;
XX-
XX- 07-OCT-1999 (first entry)
XX
XX Amino acid sequence of a Chlamydia trachomatis protein.
XX
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
XX Chlamydia trachomatis.
XX
XX WO9928475-A2.
XX
XX 10-JUN-1999.
XX
XX 27-NOV-1998; 98WO-IB01939.


```

XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
XX PN WPI; 1999-508654/42.
XX PD
XX PF Producing kinases with increased activity on nucleoside and
XX PR nucleotide analogs, used to improve conversion of prodrugs, e.g.
XX PS AZT, to active form
XX PT
XX PT Claim 27; Page 71-72; 84pp; English.
XX PS
XX CC The present sequence is the E.coli thymidylate kinase. It is a class II
XX CC (type II) kinase which generates cytotoxic nucleotide analogs from their
XX CC prodrugs, by phosphorylation, particularly the AIDS prodrug AZT.
XX CC Rational shuffling of domains from E.coli to the corresponding enzyme
XX CC from human, mouse, or yeast improves kinase activity. Increasing the
XX CC kinase activity results in higher concentrations of the active form of
XX CC the therapeutic analog (specifically AZT-triphosphate), and thus a
XX CC greater therapeutic effect, while reducing the concentration of the
XX CC toxic monophosphate intermediate. In E.coli TmpK the Arg15 in the P-loop
XX CC is absent while having basic residues in the LID region. Hence a P-loop
XX CC movement occurring upon AZT-MP binding should not have a detrimental
XX CC effect on catalysis. E.coli TmpK may have catalytic activity for
XX CC phosphorylation of AZT-monophosphate 300 times that of the wild-type
XX CC human enzyme. It has anticancer and antiviral activities.
XX SQ Sequence 210 AA;

Query Match 29.18; Score 306; DB 20; Length 210;
Best Local Similarity 40.68; Pred. No. 2.2e-24;
Matches 82; Conservative 36; Mismatches 74; Indels 10; Gaps 6;

Qy 3 KGFVLSLEGPGAGKTSVLEALLPILEKGV-EVLTTPREGVGLICEKIREVILDPSTQ 61
Db 2 kgfivieglegagkssahqsvrvlhelgigdvvtrepqgtlaeklrhlikheteep 61

Qy 62 MDKATELLLYTASRRQHLVEKVL-PALEAGKLVIMDRFIDSSVAYQFGRGLDIEADWL 120
Db 62 vtdkaellmlyaarliq-ivenvikpalmqgkwvgydrhdmsqayggggrqldphfmitl 120

Qy 121 NOFATDGLKPDLYFDIEVEEGLARIANSREVNRLDLEGLDHLKKVROGYLSLDKE 180
Db 121 ketvlgnfepdltlyldpvsvglr--argrgeldieqmdldffhrtraryielv-kd 177

Qy 181 GNRIVKIDASLPLEQVETTKA 202
Db 178 npkavvina----eqgsielvga 195

RESULT 10
AAY28786
ID AAY28786 standard; protein; 213 AA.
AC AAY28786;
XX
XX 02-NOV-1999 (first entry)
XX
XX E.coli thymidylate kinase-1.
XX
KW E.coli thymidylate kinase; E.coli TmpK; class II (type II) kinase; AIDS;
KW prodrug; cytotoxic nucleotide analog; AZT; 3'-azido-3'-deoxythymidine;
KW therapeutic analog; AZT-triphosphate; therapeutic effect; P-loop;
KW LID region; human thymidylate kinase; anticancer; antiviral.
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
XX Region 12..17
XX FT /label= P-loop
XX FT /note= "unique additional basic residues"

```

```

FT Region 138..150
XX /label= LID_region
XX
XX PN WO9941404-A2.
XX
XX PD 19-AUG-1999.
XX
XX PF 12-FEB-1999; 99WO-EP00945.
XX
XX PR 13-FEB-1998; 98EP-0102546.
XX
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX PI Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
XX PN WPI; 1999-508654/42.
XX
XX PT Producing kinases with increased activity on nucleoside and
XX PR nucleotide analogs, used to improve conversion of prodrugs, e.g.
XX PT AZT, to active form
XX
XX PS Claim 7; Page 70-71; 84pp; English.
XX
XX CC The present sequence is the E.coli thymidylate kinase. It is a class II
XX CC (type II) kinase which generates cytotoxic nucleotide analogs from their
XX CC prodrugs, by phosphorylation, particularly the AIDS prodrug AZT.
XX CC Rational shuffling of domains from E.coli to the corresponding enzyme
XX CC from human, mouse, or yeast improves kinase activity. Increasing the
XX CC kinase activity results in higher concentrations of the active form of
XX CC the therapeutic analog (specifically AZT-triphosphate), and thus a
XX CC greater therapeutic effect, while reducing the concentration of the
XX CC toxic monophosphate intermediate. In E.coli TmpK the Arg15 in the P-loop
XX CC is absent while having basic residues in the LID region. Hence a P-loop
XX CC movement occurring upon AZT-MP binding should not have a detrimental
XX CC effect on catalysis. E.coli TmpK may have catalytic activity for
XX CC phosphorylation of AZT-monophosphate 300 times that of the wild-type
XX CC human enzyme. It has anticancer and antiviral activities.
XX SQ Sequence 213 AA;

Query Match 29.08; Score 305.5; DB 20; Length 213;
Best Local Similarity 38.48; Pred. No. 2.5e-24;
Matches 78; Conservative 37; Mismatches 79; Indels 9; Gaps 6;

Qy 7 VSLLEGPGAGKTSVLEALLPILEKGV-EVLTTPREGVGLICEKIREVILDPSTQMDA- 64
Db 6 ivieglegagkttarnvvvetleqgirdmvftrepqgtlaeklrslvldiksvgdevi 65

Qy 65 --KTELLLYTASRRQHLVEKVL-PALEAGKLVIMDRFIDSSVAYQFGRGLDIEADWL 121
Db 66 tdkaevlmfyaarvq-ivetvikpalangtwigrdhldstqayggggrgidqhmldtlr 124

Qy 122 QFATDGLKPDLYFDIEVEEGLARIANSREVNRLDLEGLDHLKKVROGYLSLDKEG 181
Db 125 davlgrdfpdltyldvtpvgvlkrarq--eidrieqesfdffnrtrarylelaaqd- 181

Qy 182 NRIVKIDASLPLEQVETTKAVL 204
Db 182 ksihtidatqpleavmdairttv 204

RESULT 11
AAB72201
ID AAB72201 standard; protein; 213 AA.
XX
XX AC AAB72201;
XX
XX 04-MAY-2001 (first entry)
XX
XX E. coli thymidylate kinase amino acid sequence.
XX
XX Protein co-ordinate data; thymidylate kinase; TMK; binding pocket;

```


KW	thymidine 5' monophosphate; adenosine 5' triphosphate; ATP; TMP;
KX	antibacterial; x-ray crystal structure.
XX	Escherichia coli.
OS	
PN	WO200111025-A2.
XX	
PD	15-FEB-2001.
XX	
PF	04-AUG-2000; 2000WO-US21425.
XX	
PR	04-AUG-1999; 99US-0147117.
XX	
PA	(PHAA) PHARMACIA & UPJOHN CO.
XX	
PI	Benson TE;
XX	
DR	WPI; 2001-159976/16.
XX	
PT	New molecule/molecular complex, useful for designing inhibitors of
PT	Staphylococcus aureus thymidylate kinase (TMK), comprises a S. aureus
PT	TMK or adenosine triphosphate binding pocket .
XX	
PS	Disclosure; Page 286-287; 288pp; English.
XX	
CC	This invention relates to a molecule or molecular complex comprising a
CC	Staphylococcus aureus thymidylate kinase (TMK) thymidine 5' monophosphate
CC	(TMP) or adenosine 5' triphosphate (ATP) binding pocket. The
CC	specification includes protein co-ordinate data for thymidylate kinase.
CC	The x-ray crystal structure of S. aureus TMK is useful for solving the
CC	structure of other molecules/molecular complexes. It is also useful for
CC	designing inhibitors of S. aureus TMK activity. These inhibitors are
CC	useful for inhibiting the growth of S. aureus or causing its death. They
CC	can be used for the prevention and treatment of S. aureus TMK mediated
CC	disease and therefore exhibit antibacterial activity. Further uses
CC	include the development of chemical entities that can isomerise to short
CC	lived reaction intermediates in the chemical reaction of a substrate or
CC	other compound that interferes with S. aureus TMK. The reaction
CC	intermediates of S. aureus TMK can be deduced and can be used to design
CC	improved analogues of known S. aureus TMK inhibitors or to design new
CC	inhibitors with high specificity and stability based on the reaction
CC	intermediates. The present sequence represents the thymidylate kinase
CC	amino acid sequence from Escherichia coli.
XX	
SQ	Sequence 213 AA;
	Query Match 29.0%; Score 305.5; DB 22; Length 213;
	Best Local Similarity 38.4%; Pred. No. 2.5e-24;
	Matches 78; Conservative 37; Mismatches 79; Indels 9; Gaps 6
QY	7 VSLEGPEGAGKT ^{SV} LLEALLPILEEKGV-EVLTTRPGGVLIGEXIREVIDPSHTQMDA- 64
Db	: : :: :: : :::: : : :::
	6 ivieglegagktarnrvvettleqigdmfvtrepdggtqlae ^{klrslivldiksvgdevi} 65
QY	65 --KTELLLYIASRROHLVEKVL-PALEAGKLIVIMDRITDSSVAYQGFGRLDTIEADLNL 121
Db	-:- : -:- : :
	66 tdkaelvmfyarvq-lvetvikpalangtwvigdrhdlsctayqyggrgidqhmhtatr 124
QY	122 QFATDGLKPLDTLYFDIEVEEGRIARAANSDBRVNRLDLEGLDLHKVRVGYSLLDKREG 181
Db	: : : ::: :- : :- : :
	125 davglgfrpdltyldvtpevgkkrarg-eeldriegesdfnfnrtarylelaaqd- 181
QY	182 NRIVKIDASLPLEQVVETKA VL 204
Db	: : :
	182 kshitdatcpleavmdairttv 204
RESULT	12
AAY28792	
ID	AAY28792 standard; protein; 210 AA.
XX	
AC	AAY28792;

XX	02-NOV-1999	(first entry)	
XX	DT		
XX	DE	Thymidylate kinase-6.	
XX	DE		
KW	Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;		
KW	viral replication; DNA chain termination; AZT activation pathway;		
KW	AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;		
KW	antiviral activity; therapeutic analog.		
XX			
OS	Unidentified.		
XX			
XX	WO9941404-A2.		
XX			
XX	19-AUG-1999.		
XX			
PF	12-FEB-1999;	99WO-EP00945.	
XX			
XX	13-FEB-1998;	98EP-0102546.	
XX			
XX	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.		
XX			
PI	Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;		
XX			
XX	WPI; 1999-508654/42.		
XX			
XX	Producing kinases with increased activity on nucleoside and		
PT	nucleotide analogs, used to improve conversion of prodrugs, e.g.		
PT	AZT, to active form		
XX			
XX	Claim 7; Page 77; 84pp; English.		
XX			
CC	The present sequence is a thymidylate kinase. TmpK is involved in the		
CC	activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication		
CC	by DNA chain termination. TmpK which catalyses the second		
CC	phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate		
CC	(AZT-DP), is the rate limiting enzyme in the AZT activation pathway.		
CC	Increasing the kinase activity of TmpK results in higher concentrations		
CC	of the active form of the therapeutic analog especially AZT-triphosphate		
CC	(AZT - 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,		
CC	while reducing the concentration of the toxic monophosphate		
CC	intermediate. This has anticancer and antiviral activities.		
XX			
SQ	Sequence	210 AA;	
	Query Match	28.5%;	Score 300; DB 20; Length 210;
	Best Local Similarity	35.5%;	Pred. No. 9.4e-24;
	Matches	72; Conservative 44;	Mismatches 77; Indels 10; Gaps
QY	1	MSKGFVLSLGGPGAGKTSVLEALLPILK--GVEVLTRTPGGVLTGKREVILDP5 8	
Db	1	lmkgvfviegvgdgaktaiegfklpytkflnyqlytrepggtllaeqrllnet 60	
QY	59	HTQMDAKTELLLYIASRRHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAD 118	
Db	61	--mepiteaylfaaartehtskliipaekedqlvisdrfvfssfayqglskigidtk 117	
QY	119	WLNQAFDGLKPDLYFDIEVEEGLARIANAANSREVNRLD--LEGDLHKVKRQGYLSL 176	
Db	118	qinhairnmnpftfildcnfkealqrmqkrgnd--nlldefikgndfdtrvsyyls 175	
QY	177	LDKEGNRIVKIDASLP-LEQVVE 198	
Db	176	vdckncflingdnkqhenlekfie 198	
RESULT	13		
AAV28791			
ID	AAV28791	standard; protein; 210 AA.	
XX			
AC	AAV28791.		
XX			

RESULT	12
AAAY28792	
ID	AAAY28
XX	
AC	AAAY28

AY28792
ID AY28792 standard; protein; 210 AA.

AC AAY28792;

DT 02-NOV-1999 (first entry)
 XX Thymidylate kinase-5.
 DE
 DE
 KW Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;
 KW viral replication; DNA chain termination; AZT activation pathway;
 KW AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;
 KW antiviral activity; therapeutic analog.
 XX
 OS Unidentified.
 XX
 XX WO9941404-A2.
 PN
 XX 19-AUG-1999.
 PD
 XX 12-FEB-1999; 99WO-EP00945.
 PF
 XX 13-FEB-1998; 98EP-0102546.
 PR
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA
 XX Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
 PI
 XX WPI; 1999-508654/42.
 DR
 XX Producing kinases with increased activity on nucleoside and
 PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
 PT AZT, to active form
 XX
 XX Claim 7; Page 76; 84pp; English.
 PS
 XX The present sequence is a thymidylate kinase. TmpK is involved in the
 XX activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication
 CC by DNA chain termination. TmpK which catalyses the second
 CC phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate
 CC (AZT-DP), is the rate limiting enzyme in the AZT activation pathway.
 CC Increasing the kinase activity of TmpK results in higher concentrations
 CC of the active form of the therapeutic analog especially AZT-triphosphate
 CC (AZT = 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,
 CC while reducing the concentration of the toxic monophosphate
 CC intermediate. This has anticancer and antiviral activities.
 XX
 XX Sequence 210 AA;
 SQ

Query Match 26.3%; Score 277; DB 20; Length 210;
 Best Local Similarity 36.4%; Pred. No. 2.6e-21;
 Matches 76; Conservative 37; Mismatches 74; Indels 22; Gaps 7;
 QY 1 MSKGLVSLGEGGAGKTSVLEALLPILKEK--GVEVLTTPREGGVLGKIREVILDPSS 58
 Db 1 mkqgvfaiegvdgagktvllleafkqrfpgsflgftlfsrpggtplaeikralll--- 57
 QY 59 HTQMDAKTELLLYTASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYOGFGRGLDIEAID 118
 Db 58 heameplceaylfasarrqrlqpalqgkqvivdrfwvssyayagllkkgvldvkv 117
 QY 119 WLNQFATDGLKPDLTLYFDIEVEGLARIANSREVNRLDLEGLDHLKK-----VRQG 172
 Db 118 klnadavgsmpdftfivdcdfetalnrmakrg--qdnld-----ntvkkqadfnmtrqy 171
 QY 173 YLSLLDREGNRIVKIDASLP---LEQVVE 198
 Db 172 yhsivdnk--rvflldgnqntgcicqqfie 198

RESULT 14
 AAY28789
 ID AAY28789 standard; protein; 188 AA.
 XX
 XX AAY28789;
 XX
 XX 02-NOV-1999 (first entry)
 DT

XX Thymidylate kinase-3.
 DE
 DE
 KW Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;
 KW viral replication; DNA chain termination; AZT activation pathway;
 KW AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;
 KW antiviral activity; therapeutic analog.
 XX
 OS Unidentified.
 XX
 XX WO9941404-A2.
 PN
 XX 19-AUG-1999.
 PD
 XX 12-FEB-1999; 99WO-EP00945.
 PF
 XX 13-FEB-1998; 98EP-0102546.
 PR
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA
 XX Goody RS, Konrad M, Lavie A, Reinstein J, Schlichting I;
 PI
 XX WPI; 1999-508654/42.
 DR
 XX Producing kinases with increased activity on nucleoside and
 PT nucleotide analogs, used to improve conversion of prodrugs, e.g.
 PT AZT, to active form
 XX
 XX Claim 7; Page 74; 84pp; English.
 PS
 XX The present sequence is a thymidylate kinase. TmpK is involved in the
 XX activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication
 CC by DNA chain termination. TmpK which catalyses the second
 CC phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate
 CC (AZT-DP), is the rate limiting enzyme in the AZT activation pathway.
 CC Increasing the kinase activity of TmpK results in higher concentrations
 CC of the active form of the therapeutic analog especially AZT-triphosphate
 CC (AZT = 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,
 CC while reducing the concentration of the toxic monophosphate
 CC intermediate. This has anticancer and antiviral activities.
 XX
 XX Sequence 188 AA;
 SQ

Query Match 19.0%; Score 200; DB 20; Length 188;
 Best Local Similarity 31.0%; Pred. No. 3.3e-13;
 Matches 63; Conservative 36; Mismatches 68; Indels 36; Gaps 9;
 QY 1 MSKGLVSLGEGGAGKTSVLEALLPILKEKGVLEVLTTPREGGVLGKIREVILDPSSHT 60
 Db 1 mvdnmfivfegldgsgkttgskll-----akkmdafwtpepsnslvgkllreil--sgkt 53
 QY 61 QMDAKTELLLYTASRRQH--LVEKVLPALEAGKLVIMDRFIDSSVAYOGFGRGLDIEAID 118
 Db 54 evdnktallfaadrhentklikeelkrd-----vvcdrlyssiyaygsa-gvdenfik 108
 QY 119 WLNQFATDGLKPDLTLYFDIEVEGLARIANSREVNRLDLEGLDHLKKVRQGYLSLID 178
 Db 109 sinrya---lkpdvifllivdietaalkrvktkd-----ifekdkfkkvgdkyleae 158
 QY 179 KEGNRIVKIDASLPLEQVETTK 201
 Db 159 -eynfi-----vidttk 169

RESULT 15
 AAW98749
 ID AAW98749 standard; Protein; 191 AA.
 XX
 XX AAW98749;
 XX
 XX 31-MAR-1999 (first entry)
 DT
 XX

Search completed: February 15, 2002, 01:55:12
Job time: 2335 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:47:02 ; Search time 22.69 Seconds
(without alignments)
210.256 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKGFVLSLEGGAGKTSV.....LEQVVETTKAVLFDGMGLAK 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1053	100.0	212	4	US-09-259-109-2
2	97	9.2	265	2	US-08-898-779-2
3	97	9.2	265	2	US-08-898-779-4
4	97	9.2	265	4	US-09-224-502-2
5	97	9.2	265	4	US-09-224-502-4
6	85	8.1	260	2	US-08-879-561-8
7	85	8.1	583	4	US-09-272-414-2
8	83.5	7.9	427	3	US-09-196-857-2
9	81.5	7.7	411	4	US-09-318-443-6
10	81	7.7	800	3	US-08-776-265-3
11	79.5	7.5	369	1	US-08-844-055-2
12	79.5	7.5	369	3	US-09-006-849-2
13	79.5	7.5	429	1	US-08-906-744A-2
14	79.5	7.5	429	3	US-09-093-134-2
15	76.5	7.3	313	4	US-09-396-651B-3
16	76.5	7.3	411	4	US-09-318-443-8
17	76	7.2	389	3	US-08-972-902-4
18	75.5	7.2	1544	4	US-09-413-814-46
19	75	7.1	503	3	US-08-911-853-7
20	75	7.1	503	4	US-09-479-409-7
21	74.5	7.1	292	2	US-08-928-284-2
22	74.5	7.1	547	2	US-08-467-822-35
23	74.5	7.1	547	4	US-08-432-697-35
24	74.5	7.1	547	4	US-08-466-248-35
25	74.5	7.1	876	1	US-08-717-515-4
26	74.5	7.1	1084	1	US-08-717-515-6
27	74.5	7.1	1276	1	US-08-717-515-8

28	74	7.0	445	1	US-08-674-168-30	Sequence 30, Appl
29	74	7.0	445	3	US-08-985-908-20	Sequence 20, Appl
30	74	7.0	445	3	US-08-852-730-5	Sequence 5, Appl
31	74	7.0	445	4	US-08-985-916-12	Sequence 12, Appl
32	74	7.0	456	2	US-08-933-750C-5	Sequence 5, Appl
33	74	7.0	456	4	US-09-234-613-5	Sequence 5, Appl
34	74	7.0	540	5	PCT-US94-06362-3	Sequence 3, Appl
35	74	7.0	541	2	US-08-467-822-34	Sequence 34, Appl
36	74	7.0	541	2	US-08-447-154-19	Sequence 19, Appl
37	74	7.0	541	4	US-08-432-697-34	Sequence 34, Appl
38	74	7.0	541	4	US-08-466-248-34	Sequence 34, Appl
39	74	7.0	649	2	US-08-871-266B-16	Sequence 16, Appl
40	74	7.0	649	2	US-09-018-864A-16	Sequence 16, Appl
41	74	7.0	649	3	US-08-871-267B-22	Sequence 22, Appl
42	74	7.0	649	4	US-09-618-419-22	Sequence 22, Appl
43	74	7.0	779	1	US-08-375-134-12	Sequence 12, Appl
44	74	7.0	779	5	PCT-US95-15263-12	Sequence 12, Appl
45	73.5	7.0	415	4	US-09-318-443-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-259-109-2
; Sequence 2, Application US/09259109
; Patent No. 6270762
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Zalacaln, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Chalker, Alison F.
; APPLICANT: Ingraham, Karen A.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: tdk
; FILE REFERENCE: GM10201
; CURRENT APPLICATION NUMBER: US/09/259,109
; CURRENT FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-259-109-2

Query Match	100.0%;	Score 1053;	DB 4;	Length 212;
Best Local Similarity	100.0%;	Pred. No. 2e-106;		
-Matches 212;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSKGFVLSLEGGAGKTSVLEALLPILKEKGVLEVTREPGGVVLEKIREVILDPST 60		
Db	1	MSKGFVLSLEGGAGKTSVLEALLPILKEKGVLEVTREPGGVVLEKIREVILDPST 60		
QY	61	QMDAKTELLYIASRRHVLVEKVLPALEAGKLVIMDRFIDSSVAYQGGRLDTEADWL 120		
Db	61	QMDAKTELLYIASRRHVLVEKVLPALEAGKLVIMDRFIDSSVAYQGGRLDTEADWL 120		
QY	121	NQFATDGLKPDLTLYFFIEVEEGLARIAANSRDVNRDLLEGLDHLKKVROGYSLLDKE 180		
Db	121	NQFATDGLKPDLTLYFFIEVEEGLARIAANSRDVNRDLLEGLDHLKKVROGYSLLDKE 180		
QY	181	GNRIVKIDASLPLEQVVETTKAVLFDGMGLAK 212		
Db	181	GNRIVKIDASLPLEQVVETTKAVLFDGMGLAK 212		

RESULT 2
US-08-898-779-2
; Sequence 2, Application US/08898779
; Patent No. 5882891

QY	4	GFVLVSLEGGEGAGKTSVLEALLPILEBKGEVLTTRPGGVLTGKIKREVLDPSTQMD	63
Db	29	KVSTISIGNGCGKSTLLKALSRLLAVKEGFE-----LDGENI-----HTQ--	70
QY	64	AKTELLLYTASRRQHLVEKVLPALEAGKLVIMDFIDSSVAYQGFGR--GLDIEATDW	119
Db	71	SWKETAKKTATLIPQSS--PEVADGIGTGVGLSVSGPEPHO-----KGEGRLTATDKRRTD	122

```

RESULT      4
US-09-224-502-2
; Sequence 2, Application US/09224502
; Patent No. 6264955
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; APPLICANT: Lonetto, Michael Arthur
; APPLICANT: Warren, Patrick Vernon
; TITLE OF INVENTION: NOVEL Ferriochrome transport
; TITLE OF INVENTION: ATP-Binding Protein
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,502
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/898,779
; FILING DATE:

```

ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: P50444-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 265 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-224-502-2

Query Match 9.2%; Score 97; DB 4; Length 265;

Best Local Similarity 33.9%; Pred. No. 0.013;
Matches 40; Conservative 13; Mismatches 39; Indels 26; Gaps 6;

QY 4 GFLVLEGPAGKTSVLEALLPILEKGVLEVTTRPFGVILGKIREVILDPSTQMD 63
Db 29 GKVTIIGPNGCGKSTLLKALSRLAVKEGEV-----LDGENI-----HTQ-- 70
QY 64 AKTELLLYTASRRQHLVKVLPALAKGLVIMDRFDSSVAYQGFR--GLDIEAIDW 119
Db 71 STKEIAKKIALLPQS--PEVADGLTVGELVSYGRFPHQ----KGFGLTAEDKKKEIDW 122

RESULT 5

US-09-224-502-4
Sequence 4, Application US/09224502
Patent No. 6264955

GENERAL INFORMATION:

APPLICANT: Burnham, Martin, Karl Russel
APPLICANT: Lonetto, Michael Arthur
APPLICANT: Warren, Patrick Vernon
TITLE OF INVENTION: NOVEL Ferrichrome transport
TITLE OF INVENTION: ATP-Binding Protein
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/224,502

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/898,779

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: P50444-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 265 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear
US-09-224-502-4

Query Match 9.2%; Score 97; DB 4; Length 265;
Best Local Similarity 33.9%; Pred. No. 0.013;
Matches 40; Conservative 13; Mismatches 39; Indels 26; Gaps 6;

QY 4 GFLVLEGPAGKTSVLEALLPILEKGVLEVTTRPFGVILGKIREVILDPSTQMD 63
Db 29 GKVTIIGPNGCGKSTLLKALSRLAVKEGEV-----LDGENI-----HTQ-- 70
QY 64 AKTELLLYTASRRQHLVKVLPALAKGLVIMDRFDSSVAYQGFR--GLDIEAIDW 119
Db 71 STKEIAKKIALLPQS--PEVADGLTVGELVSYGRFPHQ----KGFGLTAEDKKKEIDW 122

RESULT 6

US-08-879-561-8
Sequence 8, Application US/08879561
Patent No. 5817482

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.

APPLICANT: Hawkins, Phillip R.

APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/879,561

FILING DATE: Herewith

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0325 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 260 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 181510

US-08-879-561-8

Query Match 8.1%; Score 85; DB 2; Length 260;

Best Local Similarity 21.6%; Pred. No. 0.25;
Matches 54; Conservative 30; Mismatches 98; Indels 68; Gaps 8;

QY 7 VSLEGPAGKTSVLEALLPILEKGV-----EVLTTREPGGVILGE 48

```
Db 24 ISIEGNTAAGKSTFVNILKQCEDEWVVPVPEVARWCNVQSTQDBFEELTMSQKNG---GN 80
Qy 49 KIREVILDPSTQMDAKTELL-----YIASRQHLVEKVLPALEAGKLVIMDRFIDSSV 103
Db 81 VLQWMEKPERWSTFTQYACLSIRAQSLNGKLKDAEKPVLFFERSVYSYRIIFASN 140
Qy 104 AYOGFGRGLDIEAI-----DWLNQFATDGLKPDLTLYFDIEVEEGLARIANS 151
Db 141 LYES-----ECNTEETWIIQDWHDMNNOFGQSLDGIIVLOATPFTCLHRIYLRG 193
Qy 152 DREVRNLDLEGLD-----LHKVKRQY-----LSLLDKEGRIYKIDASLPLE 194
Db 194 RNEEQGIPLEYLEKLYKHSHESWLLHRTLKTNFYDLQEVPIITLD-----VNEDFKKYE 247
Qy 195 QVETTKAVL 204
Db 248 SLVEKVEFL 257

RESULT 7
US-09-727-414-2
; Sequence 2, Application US/09272414A
; Patent No. 6238885
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: Histidine Kinase
; FILE REFERENCE: GM10202
; CURRENT APPLICATION NUMBER: US/09/272,414A
; CURRENT FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-727-414-2
```

```
Query Match 8.1%; Score 85; DB 4; Length 583;
Best Local Similarity 25.6%; Pred. No. 0.86; Mismatches 23; Indels 40; Gaps 7;
Matches 46; Conservative 23;

Qy 35 VLTTRPGVGLGKIREVILDPSTQMDAKTELLXIASR-RQHLVEKVLPALEAGKLV 93
Db 269 VLGINESQIILSNKMANDIM--NIDEDAKAFLLRQIEDTFKSKQTEMRDLENNAREFV 326
Qy 94 IMDRFIDSSVAYQFGRGLDIEADWLNQFATDGLKPD-----LTYFDI 138
Db 327 VTTSYIDK--TEQGGKSGVVTVTRDMTNEHNLQDMKKDFIANVSHELRTPISLQGYTES 384
Qy 139 -----EVERGLARIANSR-----EVNRLDLEGLDLHKVROGYLSLIDK 179
Db 385 IVDGIVTEPDEIKESLAWVLDESRLNRLNELLNVARMDAEGLSVNKEV-OPTAALLDK 443
```

```
RESULT 8
US-09-196-857-2
; Sequence 2, Application US/09196857A
; Patent No. 6100069
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Brown, James R.
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Throup, John P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Zhong, YiYi
; APPLICANT: Debouck, Christine
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Wang, Min
; APPLICANT: Warren, Richard L.
; APPLICANT: Schilling, Lisa K.
```

```
; TITLE OF INVENTION: No. 6100069el tig
; FILE REFERENCE: GM10118
; CURRENT APPLICATION NUMBER: US/09/196,857A
; CURRENT FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: 60/082,418
; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-196-857-2
```

```
Query Match 7.9%; Score 83.5; DB 3; Length 427;
Best Local Similarity 24.1%; Pred. No. 0.78; Mismatches 36; Indels 73; Gaps 14;
Matches 61; Conservative 36;

Qy 20 VLEALLP-----ILEEKGEV-----LTTREPG-----GVLIGE-KIREV 53
Db 67 VMNALLPNAYEAAYKEAGLEVVAQPKIDVTSMEKGQDVIAAEVVTRPEVKLGDKKNLEV 126
Qy 54 ILDPSTQMDAKTELLYIASRQHLVEKVL--PALEAGKLVIMDRFIDS--SVAYQG-- 107
Db 127 SVDVEKEVTDADVE--ERIERERNLTVELIKEAAENGDTWID-FVGSIDGVFEFGCK 183
Qy 108 ---FGRGLDI-----EADWLNQFATDGLKPDL-----LTYFDIE 139
Db 184 GENFSLGIGSQGFIPGFEDQLVGHSGAGTVDVIITFFEDYQADLAGKEAFVTTIHEVK 243
Qy 140 VEEGLA---RIAANSREVRNRLDLEGLDLHKVKRQGYLSLID--KEGNRIYKIDASLP 193
Db 244 AKEVPALDDELAKDIDEVETL----ADLKEKYKELAAAKEAYKDAVEGAADTAVEN 299
Qy 194 EQVETTKAVLFD 206
Db 300 AEIVELPEEMIHE 312
```

```
RESULT 9
US-09-318-443-6
; Sequence 6, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AII1, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-318-443-6
```

```
Query Match 7.7%; Score 81.5; DB 4; Length 411;
Best Local Similarity 23.4%; Pred. No. 1.2; Mismatches 35; Indels 75; Gaps 12;
Matches 57; Conservative 35;

Qy 3 KGLVLSLEGPEGAGKT-----SVLEAL-LPILKEKGEVLTRE----- 40
Db 74 KGRDVIASQSGTGCTATFSISVLQCLDIQVRETQALLIAPTRELAVQIQKGLLAGDYM 133
Qy 41 -----PGGVLIGEKIREVILDPSTQMDAKTELLYIASRQHLVEKVLPALE---BAG 90
Db 134 NVOCHACTIGTGVNGEDIRK--LDYQG-HVVAGTPCRVDIMIRRSRLRTRAKMLVLDEAD 190
```


QY 91 KLV-----IMD--REFIDSSVAYQFGRLDIEADLWLNQFATDGLKPDLTLYFDIEVE 141
 Db 191 EMLNKGKEQIYDVRPLPQATQVVLISATLPHLEILNTKFTDPI----- 237
 QY 142 EGLARIAANSREYNRLDLEGLDHLKVKVROGYLSLLDKEGNRIVKIDASLPLEQVETTK 201
 Db 238 ----RILVKRD---ELTLEG-----IKOFFVAVEREW----KFDTLCDLYDTLITQ 279
 QY 202 AVL F 205
 Db 280 AVIF 283

RESULT 10
 US-08-776-265-3
 ; Sequence 3, Application US/08776265
 ; Patent No. 6001631
 ; GENERAL INFORMATION:
 ; APPLICANT: BLANCHE, Francis
 ; APPLICANT: CAMERON, Beatrice
 ; APPLICANT: CROUZET, Joel
 ; APPLICANT: FAMECHON, Alain
 ; APPLICANT: FERRERO, Lucia
 ; TITLE OF INVENTION: No. 6001631el Topoisomerase IV, Corresponding
 ; TITLE OF INVENTION: Nucleotide Sequences and Uses Thereof
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 ; STREET: 1300 I. Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/776,265
 ; FILING DATE: 24-JAN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einaudi, Carol P.
 ; REGISTRATION NUMBER: 32,220
 ; REFERENCE/DOCKET NUMBER: 03806.0394-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 408-4000
 ; TELEFAX: (202) 408-4444
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 800 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-776-265-3

Query Match 7.7%; Score 81; DB 3; Length 800;
 Best Local Similarity 28.1%; Pred. No. 3.8;
 Matches 48; Conservative 29; Mismatches 62; Indels 32; Gaps 9;
 QY 50 IREVILDPSTQMDA---KTELLLYIASRRQHLVEKVLPALEACKLVIMDRFD---SSV 103
 Db 343 IRIIDSYLNHQIEVAVNRRTKFDLNAEKRMHIVEGLIKALS-----ILDKVIELRSK 397
 QY 104 AYQFGRLDIEADLWLNQFATDGLKPDLTLY----FDIEVEEG-----LARIA 148
 Db 398 NKRDAKENL-IEVEFTEQAEALVM--LQLYRTNTDIVALEGEHKEALIKQLRHIL 454
 QY 149 ANSDREYNRLDLEGLDHLKVKVROGYLSLLDKEGNRIVKID--ASLPLEQVV 197

Db 455 DNHALLNVIKEELNEIKKFKSERLSLIEAEIEEI-KIDKEVWVPSEVI 504
 RESULT 11
 US-08-844-055-2
 ; Sequence 2, Application US/08844055
 ; Patent No. 5747313
 ; GENERAL INFORMATION:
 ; APPLICANT: Lawlor, Elizabeth
 ; TITLE OF INVENTION: No. 5747313el Compounds
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road
 ; CITY: King of Prussia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406-0939
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/844,055
 ; FILING DATE: 18-APR-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 9607993.4
 ; FILING DATE: 18-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gimmi, Edward R.
 ; REGISTRATION NUMBER: 38,891
 ; REFERENCE/DOCKET NUMBER: P31457-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-270-4478
 ; TELEFAX: 610-270-5090
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 369 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-844-055-2

Query Match 7.5%; Score 79.5; DB 1; Length 369;
 Best Local Similarity 20.1%; Pred. No. 1.7;
 Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
 QY 6 LVSLEGPE--GAGKTSVLEALLPI-----LEEKGVETLTREPGGVLTGE---K 49
 Db 165 LNTLGNPESRAAYRQALIDVLTPLKETLSKDSQRLEENPLRVLDSEKEDKVAVENAPS 224
 QY 50 IREVILDPSTQMDAKTELL--YIASRR-----QHLVEKVLPALEACKLVIM- 95
 Db 225 ILDFLDEESQTHFDVAVSQMLENLGVYDIIDTNMVRGLDYNNHTIFETIEGNDLTVCA 284
 QY 96 -DRFIDSSVAY-----QGFGRLDIEADLWLNQFATDGLKPDLTLYFDIEVEEGLARIA 148
 Db 285 GGRY-DGLVAYFGGPGTAGFGGLGVERL----- 312
 QY 149 ANSDREYNRLDLEGLDHLKVKVROGYLSLLDKEGNRIVKIDASLPLEQVETTKAVLFDG 207
 Db 313 -----LLILEKQG-----VALPIENALDVIYIAVLGDG 339
 RESULT 12
 US-09-006-849-2
 ; Sequence 2, Application US/09006849

Patent No. 6071731
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6071731e1 Compounds
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,849
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-849-2

Query Match 7.5%; Score 79.5; DB 3; Length 369;
Best Local Similarity 20.1%; Pred. No. 1.7;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
Qy 6 LVSLGGP--GAGKTSVLEALLPI-----LEEKGEVLTTRPGGVGLIGE---K 49
Db 165 LNTLGNPESRAAYRQALIDYLTPLKETSQSRRLNPLRVLDSEKEDKVAVENAPS 224
Qy 50 IREVLDPSTQMDAKTELL-----YIASRR-----OHLVEKVLPALEAGKLVIM- 95
Db 225 ILDFDEESQTHFDVAVSQMLNGLVDYIIDTNMVRGLDYNNHTIFEITEGNDLTVCA 284
Qy 96 -DRFIDSSVAY-----QGFRGLDIEADWLNOFATDGLKPDLTLYFDIEVEEGLARIA 148
Db 285 GGRY-DGLVAYFGGPETAGFGGLGVERL-----LILEKQG----- 312
Qy 149 ANSDREVNRLDGLDLHKYRQGYLSLLDKEGNRIVKIDASLPLEQVVTTKAVLFDG 207
Db 313 -----LILEKQG-----VALPIENALDVYIAVLGDG 339

RESULT 13
US-08-906-744A-2
Sequence 2, Application US/08906744A
Patent No. 5795758
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Greenwood, Rebecca
APPLICANT: Lawlor, Elizabeth

TITLE OF INVENTION: NOVEL HISS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,744A
FILING DATE: 06-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-906-744A-2

Query Match 7.5%; Score 79.5; DB 1; Length 429;
Best Local Similarity 20.1%; Pred. No. 2.1;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;
Qy 6 LVSLGGP--GAGKTSVLEALLPI-----LEEKGEVLTTRPGGVGLIGE---K 49
Db 165 LNTLGNPESRAAYRQALIDYLTPLKETSQSRRLNPLRVLDSEKEDKVAVENAPS 224
Qy 50 IREVLDPSTQMDAKTELL-----YIASRR-----OHLVEKVLPALEAGKLVIM- 95
Db 225 ILDFDEESQTHFDVAVSQMLNGLVDYIIDTNMVRGLDYNNHTIFEITEGNDLTVCA 284
Qy 96 -DRFIDSSVAY-----QGFRGLDIEADWLNOFATDGLKPDLTLYFDIEVEEGLARIA 148
Db 285 GGRY-DGLVAYFGGPETAGFGGLGVERL-----LILEKQG----- 312
Qy 149 ANSDREVNRLDGLDLHKYRQGYLSLLDKEGNRIVKIDASLPLEQVVTTKAVLFDG 207
Db 313 -----LILEKQG-----VALPIENALDVYIAVLGDG 339

RESULT 14
US-09-093-134-2
Sequence 2, Application US/09093134
Patent No. 6040162
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Greenwood, Rebecca
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: NOVEL HISS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/093,134
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/906,744
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,055
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-1/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

```

Query Match          7.5%; Score 79.5; DB 3; Length 429;
Best Local Similarity 20.1%; Pred. No. 2.1;
Matches 48; Conservative 32; Mismatches 58; Indels 101; Gaps 10;

QY      6 LVSLEGPE--GAGKTSVLEALLPT-----LLEKGEVLTTRPGGVLLGE---K 49
      | : | | : : : : : | : | | : | : | : |
Db      165 LNTLGNPESRAAYQALIDYLTPLKETLSKDSQRLEENPLRVLDSKEKEKDVAVENAPS 224

QY      50 IREVLPDSHTQMDAKTELL-----YIASRR-----QHLVEKVLPALEAGKLVIM- 95
      | : | | : | : | : | | : | : | : | : |
Db      225 ILDFLDEESQTHFDVAVSQMLENLGVWDYIIDTNMVRGLDYNNHTIFEITEGNDLTVCA 284

QY      96 -DRFIDSSVAY-----QGFGRGDIIRALDNLQFATDGLKPDLTLYFDIEVEEGLARIA 148
      | : | | | | : | : | : | : | : | : |
Db      285 GGXY-DGLVAYFGGFTAGFGGLGVERL----- 312

QY      149 ANDREVNRLDLEGLDHKKVRQGYLSLLDKEGNRIVKIDASLPLEQOVETTKAVLFDG 207
      | : | : | : | : | : | : | : | : | : |
Db      313 -----LILTEKQG-----VALPTENALDVIYIIVLGDG 339

```

```

RESULT 15
US-09-396-651B-3
; Sequence 3, Application US/09396651B
; Patent No. 6225076
; GENERAL INFORMATION:
; APPLICANT: Darst, Seth A
; APPLICANT: Zhang, Gongyi
; APPLICANT: Campbell, Elizabeth
; APPLICANT: Minakin, Leonid
; APPLICANT: Severinov, Konstantin
; TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS

```

```

; TITLE OF INVENTION: OF USE THEREOF
; FILE REFERENCE: 600-1-258
; CURRENT APPLICATION NUMBER: US/09/396,651B
; CURRENT FILING DATE: 1999-09-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Thermus aquaticus
US-09-396-651B-3

```

Query Match	7.38;	Score	76.5;	DB	4;	Length	313;
Best Local Similarity	24.2%;	Pred. No.	2.8;				
Matches	47;	Conservative	35;	Mismatches	75;	Indels	37;
Gaps	10;						
QY	32	GVEVLTREPT-GVYLIGEKIREVILD--PSHTQMDAKTELLY-----IASRRQHLVEKVL	84				
Db	21	GEFVLEPLERGFVTLGNPLRILLSSIPGTAVTSVYIEDVLHFEFTGPKVEDVVEIL	80				
QY	85	PALEAGKLVIMDRFD-----SSVAYQGEFGRGLDTEADWLNAQFATDGLKPDL-----	132				
Db	81	NLKE---LVV--REFDPRWRTLILRAEG-PKEVRAVDFTPSADVEIMNPDLHATLEEG	134				
QY	133	-TLFYDTEVEBGLARIANSDRVNRILDLEGLD-LHKVKRO-----GYLSLLDKE	180				
Db	135	GKLYMEVVRDREGVGVPAERGIKDRINAIPVDALFSPVRVAFQVETRILGQRTDLQKL	194				
QY	181	GNRIVKIDASLPLE	194				
Db	195	TLRIWTDGTSVTPLE	208				

Search completed: February 15, 2002, 01:55:56
Job time: 534 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:50:02 ; Search time 33.21 Seconds
(without alignments)
486.269 Million cell updates/sec

Title: US-09-749-972-2
Perfect score: 1053
Sequence: 1 MSKGLVSLGPEGAGKTSV.....LEQVVTTKAVLFDGMGLAK 212
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	659	62.6	211	2 F86674	thymidylate kinase
2	551	52.3	210	2 B83655	thymidylate kinase
3	528	50.1	212	2 S66058	thymidylate kinase
4	404	38.4	210	2 E83275	thymidylate kinase
5	392.5	37.3	195	2 H70383	thymidylate kinase
6	387.5	36.8	206	2 D75560	thymidylate kinase
7	373	35.4	1089	2 T36663	protein kinase, tr
8	368.5	35.0	197	2 E72294	thymidylate kinase
9	361	34.3	203	2 E71546	probable thymidyla
10	361	34.3	203	2 G81700	thymidylate kinase
11	357.5	34.0	206	2 H81932	probable dtmP kina
12	356	33.8	206	2 H81173	thymidylate kinase
13	329.5	31.3	206	2 A86525	thymidylate kinase
14	329.5	31.3	206	2 F72098	thymidylate kinase
15	314	29.8	212	2 A82128	thymidylate kinase
16	307.5	29.2	205	2 A75165	thymidylate kinase
17	306	29.1	237	2 A64153	dtmP kinase (EC 2.
18	305.5	29.0	213	2 G64853	dtmP kinase (EC 2.
19	304.5	28.9	213	2 H85672	thymidylate kinase
20	300	28.5	210	2 F64200	thymidylate kinase
21	294.5	28.0	196	2 E69257	thymidylate kinase
22	294	27.9	205	2 A71177	thymidylate kinase
23	289.5	27.5	203	2 G71674	probable thymidyla
24	280	26.6	210	2 S73474	thymidylate kinase
25	262	24.9	212	2 A84971	dtmP kinase (EC 2.
26	262	24.9	233	2 H82944	thymidylate kinase
27	238.5	22.6	199	2 H84343	thymidylate kinase
28	204.5	19.4	217	2 F82788	thymidylate kinase
29	201	19.1	98	2 S48605	hypothetical prote

30	200	19.0	188	2 F64336	dtmP kinase (EC 2.
31	198.5	18.9	191	2 B64704	thymidylate kinase
32	195.5	18.6	191	2 B71814	thymidylate kinase
33	185.5	17.6	216	1 K1BY78	dtmP kinase (EC 2.
34	174	16.5	208	2 H71333	probable thymidyla
35	172	16.3	194	2 F69102	thymidylate kinase
36	170.5	16.2	192	2 F81347	thymidylate kinase
37	169	16.0	208	2 E72514	probable dtmP kina
38	154	14.6	236	2 S76121	hypothetical prote
39	145	13.8	257	2 H70198	thymidylate kinase
40	133.5	12.7	252	2 F65084	hypothetical prote
41	131.5	12.5	230	2 H65084	hypothetical prote
42	129	12.3	188	2 T52029	dtmP kinase (EC 2.
43	125.5	11.9	210	2 T41553	thymidylate kinase
44	123	11.7	211	2 S26845	dtmP kinase (EC 2.
45	115.5	11.0	218	2 T24244	hypothetical prote

ALIGNMENTS

RESULT 1

F86674
thymidylate kinase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: F86674
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.;
Genome Res. in press, 2001
A;Title: The complete genome sequence of the lactic acid bacterium.
A;Reference number: A86625
A;Accession: F86674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <STO>
A;Cross-references: GB:AE005176; NID:gl2723271; PIDN:AAK04496.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yeab
C;Superfamily: dtmP kinase

Query Match 62.6%; Score 659; DB 2; Length 211;
Best Local Similarity 61.7%; Pred. No. 2e-42;
Matches 124; Conservative 39; Mismatches 38; Indels 0; Gaps 0;

Qy	4	GFLVSLGPEGAGKTSVLEALLPILEKGVVLTTPREGVLTGKIREVLDPSTQMD	63
Db	3	GILSLGPDGAGKTTVLKEILPEIQMKREIVPTREGVVRVAEIRQIILDPKNTDID	62
Qy	64	AKTELLLYIASRROHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEAIDMLAQF	123
Db	63	SKTEMLFAAARRLHMOKRMLPALQAGKVIIVDRFIDSSVAYQGFGRGLVEVDMLNYF	122
Qy	124	ATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLBGLDHHKVRQGLSLDKEGKR	183
Db	123	ATDGLKPDLTLYFDVDIDVALERIMKNRADEVNRLDLERAEMHRKVRGYLEIVVKEPER	182
Qy	184	IVKIDASLPLEQVVTETTKAVL	204
Db	183	FKVIDASQPLEKVVADTILSVL	203

RESULT 2

B83655
thymidylate kinase tmk [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C;Accession: B83655
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A;Reference number: A83650; MUID:20263314

A:Accession: B93655
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AP001507; GB:BA000004; NID:gi0172612; PIDN:BA03761.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: tmk
C:Superfamily: dTMP kinase

Query Match 52.3%; Score 551; DB 2; Length 210;
Best Local Similarity 54.8%; Pred. No. 2.4e-34;
Matches 108; Conservative 38; Mismatches 51; Indels 0; Gaps 0;
Qy 1 MSKGLVSLGEGGAGKTSVLEALPILKEKGVETLTTRPGGVILGKIREVILDPST 60
Db 1 MTKGCFIVGEGGAGKTSALDAIEMLENGLSVVRTPGPIPIAQIRSIILDVHT 60
Qy 61 QMDATELLYIASRQHLVKEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIAIDWL 120
Db 61 RMDPRTEALLYAARRQHLVKEKVLPALEAGHVLCDFIDSSVAYQGFGRGLDIAI 120
Qy 121 NQFATDGLKPDLTLYFDIEVEEGLARIAANSRVNRLDLEGLDHLKKVROGYLSLLKE 180
Db 121 NEFAIEGRYPDLTLLFRYDPDVGSLRIHQDSREQNRLDQELTFHQVKEGYRIVET 180
Qy 181 GNRIKVIDASLPLEQV 197
Db 181 PERVVEIDANQSFQV 197

RESULT 3
S66058
thymidylate kinase tmk - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jan-2001
C:Accession: S66058; D69724
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
A:Reference number: S65967; MUID:96051385
A:Accession: S66058
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <OGA>
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05264.1; PID:g467418
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertel
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabet, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, K.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: D69724
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <KUN>
A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11804.1; PID:g2632295
A:Experimental source: strain 168
C:Genetics:
A:Gene: tmk
C:Superfamily: dTMP kinase

C:Keywords: nucleotide binding; P-loop
F:10-17/Region: nucleotide-binding motif A (P-loop)

Query Match 50.1%; Score 528; DB 2; Length 212;
Best Local Similarity 51.7%; Pred. No. 1.3e-32;
Matches 106; Conservative 37; Mismatches 62; Indels 0; Gaps 0;
Qy 4 GFLVSLGEGGAGKTSVLEALPILKEKGVETLTTRPGGVILGKIREVILDPSTQMD 63
Db 3 GLFITFEGGAGKTTVLOEIKNLTAEGLQVMATREPGGIDIAEQIREVILNENILMD 62
Qy 64 AKTELLYIASRQHLVKEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIAIDWLNOF 123
Db 63 PKTEALLYAARRQHLVKEKVLPALEAGKLVLCDFIDSSVAYQGFGRGLDIAIDVLSINEF 122
Qy 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRVNRLDLEGLDHLKKVROGYLSLLKEGNNR 183
Db 123 AIGDMMPHVTYVFSIDPEGLKRIYANGSREKNRLDLEKLPDFTTKVQEGYQELMKRFFER 182
Qy 184 IVKIDASLPLEQVETTKRAVLFDGM 208
Db 183 PHSVDAGOSKDLVQDVLKVIDEAL 207

RESULT 4
E83275
thymidylate kinase PA2962 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83275
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: AB2950; MUID:20437337
A:Accession: E83275
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AE004722; GB:AE004091; NID:g9949054; PIDN:AAG06350.1; GSPDB:G
A:Experimental source: strain PA01
C:Genetics:
A:Gene: tmk; PA2962
C:Superfamily: dTMP kinase

Query Match 38.4%; Score 404; DB 2; Length 210;
Best Local Similarity 47.7%; Pred. No. 2.4e-23;
Matches 92; Conservative 25; Mismatches 74; Indels 2; Gaps 1;

Qy 4 GFLVSLGEGGAGKTSVLEALPILKEKGVETLTTRPGGVILGKIREVILDPSTQMD 63
Db 3 GLFITFEGGAGKTSNDRYLAERLREGRGIEVQLTRPGGTPLAERIPELLAPSDEPM 62
Qy 64 AKTELLYIASRQHLVKEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIAIDWLNOF 123
Db 63 ADTELLMFAARQHLVIRPALARGAVLCDFIDSSVAYQGFGRGLPEARIALESF 122
Qy 124 ATDGLKPDLTLYFDIEVEEGLARIAANSRVNRLDLEGLDHLKKVROGYLSLLKEGNNR 183
Db 123 VQGLRDLPLTLVFDLPVETGLARAARG--RLDRPEQEDRRRFFAVRQTYLQRAAQAPER 180
Qy 184 IVKIDASLPLEQV 196
Db 181 YQVLDAGLPLAEV 193

RESULT 5
H70383
thymidylate kinase - Aquifex aeolicus
C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C:Accession: H70383
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, V.

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666

A:Accession: H70383

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-195 <AQF>

A:Cross-references: GB:AE000716; NID:g2983478; PIDN:AAC07063.1; PID:g2983484; GB:AE00065

A:Experimental source: strain VF5

C:Genetics: tmk

C:Superfamily: dTMP kinase

Query Match

Best Local Similarity 37.3%; Score 392.5; DB 2; Length 195;

Matches 87; Conservative 42; Mismatches 51; Indels 11; Gaps 4;

Y 6 LVSLGPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLTGKIREVILDPSTOMDAK 65

Y 2 LIAFEGIDGSKTQAKLYEYKQGYFVSLYRPGGKVGVEVLEILT---EELDER 58

Y 66 TELLTYASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNOFAT 125

Y 59 TELLFEASRSKLIIEKIIPDLARKDVILDFVLTIAQGYGKGLDVEFIKLNFEAT 118

Y 126 DGLKPDLTLYDFIEVEGLARIAANSDEVNRDLDEGLDLHKVKRQGYLSLDEKGNRIY 185

Y 119 RGVKPDITLLDIPVDIALRRL-----KEKNR--FENKEFLEKVRKGFLEAKEEN-VV 170

Y 186 KIDASLPLEQV 196

Y 171 VIDASGEEEV 181

RESULT 6

D75560

thymidylate kinase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: D75560

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: D75560

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <WHI>

A:Cross-references: GB:AE001873; GB:AE000513; NID:g6457764; PIDN:AAF09698.1; PID:g645777

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0111

A:Map position: 1

C:Superfamily: dTMP kinase

Query Match

Best Local Similarity 36.8%; Score 387.5; DB 2; Length 206;

Matches 85; Conservative 36; Mismatches 72; Indels 3; Gaps 2;

Y 1 MSKGLVSLGPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLTGKIREVILDPST 60

Y 1 MSOGLFITLGGPEGAKTQLARLEARLRAAGHAVTVTRPGGTPLGTRVREVVLPDPA-V 59

Y 61 QMDAKTELLTYASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWL 120

Y 119 RGVKPDITLLDIPVDIALRRL-----KEKNR--FENKEFLEKVRKGFLEAKEEN-VV 170

Y 186 KIDASLPLEQV 196

Y 171 VIDASGEEEV 181

RESULT 6

D75560

thymidylate kinase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: D75560

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: D75560

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <WHI>

A:Cross-references: GB:AE001873; GB:AE000513; NID:g6457764; PIDN:AAF09698.1; PID:g645777

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0111

A:Map position: 1

C:Superfamily: dTMP kinase

Query Match

Best Local Similarity 43.4%; Score 387.5; DB 2; Length 206;

Matches 85; Conservative 36; Mismatches 72; Indels 3; Gaps 2;

Y 1 MSKGLVSLGPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLTGKIREVILDPST 60

Y 1 MSOGLFITLGGPEGAKTQLARLEARLRAAGHAVTVTRPGGTPLGTRVREVVLPDPA-V 59

Y 61 QMDAKTELLTYASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWL 120

Y 119 RGVKPDITLLDIPVDIALRRL-----KEKNR--FENKEFLEKVRKGFLEAKEEN-VV 170

Y 186 KIDASLPLEQV 196

Y 171 VIDASGEEEV 181

RESULT 6

D75560

thymidylate kinase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: D75560

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: D75560

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <WHI>

A:Cross-references: GB:AE001873; GB:AE000513; NID:g6457764; PIDN:AAF09698.1; PID:g645777

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0111

A:Map position: 1

C:Superfamily: dTMP kinase

Db 60 EIEPLGEFLYSASRAQLVREVLRPALERGETVLCDRYADSSLAYQAGRGSLPLLRQI 119

QY 121 NQFATDGLKPDLTLYDFIEVEGLARIAANSDEVNRDLDEGLDLHKVKRQGYLSLDEK 180

Db 120 TAEVTGSLTGLIVLLDLDPALGLQRAARG--QPDRLAQADLTFFHRRVQGFLLAHAE 177

QY 181 GNRIVKIDASLPLEQV 196

Db 178 PQRFVLDAATRPEDL 193

RESULT 7

T36663

protein kinase, transmembrane - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36663

R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, submitted to the EMBL Data Library, March 1999

A:Reference number: Z21611

A:Accession: T36663

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1089 <SEE>

A:Cross-references: EMBL:AL035636; PIDN:CAB38479.1; GSPDB:GN00070; SCOEDB:SCH5.05c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCH5.05c

Query Match 35.4%; Score 373; DB 2; Length 1089;

Best Local Similarity 40.9%; Pred. No. 4e-20;

Matches 81; Conservative 35; Mismatches 78; Indels 4; Gaps 1;

QY 4 GFLVSLGPEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLTGKIREVILDPSTQMD 63

Db 509 GFFIALEGGDGAGKSTQAEALAEWIRKGHEVLTREPQATPVGKRLRSILLDVSSAGLS 568

QY 64 AKTELLTYASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNOF 123

Db 569 HRAELLYAADRAEHVDVTVVRPALERGAVVVSORIDSSVAYQAGRDLSFTETARINRW 628

QY 124 ATDGLKPDLTLYDFIEVEGLARIAANSDEVNRDLDEGLDLHKVKRQGYLSLDEKGNR 183

Db 629 ATNGVLPHLVLLDVAPEARERTEAPD---RLESEPAEFHARVRSGLTAAADPGR 684

QY 184 IVKIDASLPLEQVETTK 201

Db 685 YLVVDAGQEPEAVTTVVR 702

RESULT 8

E72294

thymidylate kinase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: E72294

R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H. Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom

A:Reference number: A72200; MUID:99287316

A:Accession: E72294

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-197 <ARN>

A:Cross-references: GB:AE001769; GB:AE000512; NID:g4981639; PIDN:AND36175.1; PID:g4

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1099

C:Superfamily: dTMP kinase

Query Match 35.0%; Score 368.5; DB 2; Length 197;
Best Local Similarity 42.6%; Pred. No. 1e-20;
Matches 81; Conservative 39; Mismatches 59; Indels 11; Gaps 4;

Qy 7 VSLEGPAGKTSVLEALLPILKEKGVLEVTTRPGGVLIGEKIREVILDPSTQMDAKT 66
Db 3 ITFEGIDSGKSTQIQLLAQYLERKGGKVIKREPGETGEKIRKILLE---EETPKA 59

Qy 67 ELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNQFATD 126
Db 60 ELFLFLASRNL-LVTEIKQYLSEGVALLDRYTDSSVAYQGFGRNLGKEIVEELNDFATD 118

Qy 127 GLKPDLTLYFDIEVEGLARIAANSDRVNRDLLEGDLHKKVRQYLSLDDKEGNRIYK 186
Db 119 GLIPDLTYIDVDETALKR-----KGELNR--FEKREFLERVREGVLVAREHPRIW 171

Qy 187 IDASLPLEQV 196
Db 172 LQGRSIEEI 181

RESULT 9
E71546
probable thymidylate kinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: E71546
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:99000809
A:Accession: E71546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <ARN>
A:Cross-references: GB:AE001292; GB:AE001273; NID:g3328586; PIDN:AAC67780.1; PID:g332859
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: tdk
C:Superfamily: dTMP kinase

Query Match 34.3%; Score 361; DB 2; Length 203;
Best Local Similarity 40.9%; Pred. No. 3.9e-20;
Matches 83; Conservative 40; Mismatches 60; Indels 20; Gaps 6;

Qy 5 FLVLEGPAGKTSVLEALLPILKEKGVLEVTTRPGGVLIGEKIREVILDPSTQMDA 64
Db 2 FIV-VEGEGAGKTQFTQALSKRLIEGKEIVTTREPCCSLGDSVRGLLDPEQ-KISP 59

Qy 65 KTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNOFA 124
Db 60 YAEALLFLAARQHIQEKILPALESGKTVICDRFHDSTIVYQIAGGLG-----EAFV 112

Qy 125 TD-----GLKPDLTLYFDIEVEGLARIAANSDRVNRDLLEGDLHKKVRQYLS 175
Db 113 TDLCYRVVGDEPFLPDITFLDLPEKEGLLR--KTRQNLDRFEQKPTSFHRAAREGFTS 170

Qy 176 LLDKEGNRIKIDASLPLEQVVE 198
Db 171 LAERAPORYKVLDPSTQMDAKT 193

RESULT 10
G81700
thymidylate kinase TC0460 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: G81700
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3
A:Reference number: A81500; MUID:20150255
A:Accession: G81700
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <TET>
A:Cross-references: GB:AE002314; GB:AE002160; NID:g7190495; PIDN:AAF39311.1; PID:g7191
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0460
C:Superfamily: dTMP kinase

Query Match 34.3%; Score 361; DB 2; Length 203;
Best Local Similarity 39.9%; Pred. No. 3.9e-20;
Matches 81; Conservative 40; Mismatches 62; Indels 20; Gaps 5;

Qy 5 FLVLEGPAGKTSVLEALLPILKEKGVLEVTTRPGGVLIGEKIREVILDPSTQMDA 64
Db 2 FIV-VEGEGAGKTQFTQALSKRLIEGKEIVLTREPCCSALGELQRLDLVDVTQ-EISS 59

Qy 65 KTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNOFA 124
Db 60 YAEALLFLAARQHIQEKILPALESGKTVICDRFHDSTIVYQIAGGLG-----EAFV 112

Qy 125 TD-----GLKPDLTLYFDIEVEGLARIAANSDRVNRDLLEGDLHKKVRQYLS 175
Db 113 TDLCYRVVGDEPFLPDITFLDLPEKEGLLR--KTRQNLDRFEQKPTSFHRAAREGFTS 170

Qy 176 LLDKEGNRIKIDASLPLEQVVE 198
Db 171 LAERAPORYKVLDPSTQMDAKT 193

RESULT 11
H81932
probable dTMP kinase (EC 2.7.4.9) NMA0869 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81932
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 2249
A:Reference number: A81775; MUID:20222556
A:Accession: H81932
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84149.1; PID:g73
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: tmk; NMA0869
C:Superfamily: dTMP kinase
C:Keywords: phosphotransferase

Query Match 34.0%; Score 357.5; DB 2; Length 206;
Best Local Similarity 39.9%; Pred. No. 7.2e-20;
Matches 79; Conservative 41; Mismatches 75; Indels 3; Gaps 2;

Qy 7 VSLEGPAGKTSVLEALLPILKEKGVLEVTTRPGGVLIGEKIREVILDPSTQMDAKT 66
Db 6 ITLDIGAGKSTNLAVIKAWFERGLPVLTREPCCGTPVGEALREILNP-ETKAGURA 64

Qy 67 ELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADWLNQFATD 126
Db 65 ELLMFAARMOHIEDVILPALSDGHIHVVSORFTATATAYQGGGRMPSEDEILEHWWQV 124

Qy 127 GLKPDLTLYFDIEVEGLARIAANSDRVNRDLLEGDLHKKVRQYLSLDDKEGNRIYK 186
Db 125 GLRPDLTLLDVPLEYSNARI--QOTREKDRFEQEQAFMRVRSVRLNRAAACPERYAV 182

Search completed: February 15, 2002, 01:56:50
Job time: 408 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:55:57 ; Search time 24.29 Seconds
(without alignments)
320.006 Million cell updates/sec

Title: US-09-749-972-2

Perfect score: 1053

Sequence: 1 MSKGLVSLGPEGAGKTSV.....LEQVETTKAVLFDGMGLAK 212

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	659	62.6	211	1 KTHY_LACLA	Q9ci94 lactococcus
2	551	52.3	210	1 KTHY_BACHD	Q9kg19 bacillus ha
3	528	50.1	212	1 KTHY_BACSU	P37537 bacillus su
4	404	38.4	210	1 KTHY_PSEAE	Q9hzn8 pseudomonas
5	392.5	37.3	195	1 KTHY_AQUAE	Q67099 aquifex aeo
6	387.5	36.8	206	1 KTHY_DEIRA	Q9ry40 deinococcus
7	368.5	35.0	197	1 KTHY_THEMA	Q9x013 thermotoga
8	361	34.3	203	1 KTHY_CHLMO	Q9pkk5 chlamydia m
9	361	34.3	203	1 KTHY_CHLTR	Q84191 chlamydia t
10	357.5	34.0	206	1 KTHY_NEIMA	Q9jve7 neisseria m
11	356	33.8	206	1 KTHY_NEIMB	Q9k0d9 neisseria m
12	337	32.0	208	1 KTHY_CAUCR	Q9rq19 caulobacter
13	329.5	31.3	206	1 KTHY_CHLPN	Q9z8f5 chlamydia p
14	314	29.8	209	1 KTHY_PASMO	Q9cke9 pasteurella
15	314	29.8	212	1 KTHY_VIECH	Q9kq12 vibrio chol
16	307.5	29.2	205	1 KTHY_PYRAB	Q9v1e9 pyrococcus
17	306	29.1	210	1 KTHY_HABIN	P44719 haemophilus
18	305.5	29.0	213	1 KTHY_ECOLI	P37345 escherichia
19	305	29.0	212	1 KTHY_YERPE	Q9g169 yersinia pe
20	300	28.5	210	1 KTHY_MYCGE	P47252 mycoplasma
21	294.5	28.0	196	1 KTHY_ARCFU	Q30175 archaeoglob
22	294	27.9	205	1 KTHY_PYRHO	Q59366 pyrococcus
23	289.5	27.5	203	1 KTHY_RICPR	Q9zcn9 rickettsia
24	280	26.6	210	1 KTHY_MYCPN	P75106 mycoplasma
25	262	24.9	212	1 KTHY_BUCAI	P57434 brucella ap
26	262	24.9	230	1 KTHY_UREPA	Q9prc5 ureaplasma
27	238.5	22.6	199	1 KTHY_HALN1	Q9hmv4 halobacteri
28	200	19.0	188	1 KTHY_METJA	Q57741 methanococ
29	198.5	18.9	191	1 KTHY_HELPY	Q26009 helicobacte
30	198.5	18.9	208	1 KTHY_XYLFA	Q9pf57 xyella fas
31	195.5	18.6	191	1 KTHY_HELPJ	Q9zjes helicobacte
32	188	17.9	189	1 KTHI_SULSO	Q9uxg7 sulfolobus
33	185.5	17.6	216	1 KTHY_YEAST	P00572 saccharomyc

34 172 16.3 194 1 KTHY_METHH 027793 methanobact
35 170.5 16.2 192 1 KTHY_CAMJE Q9ppf3 campylobact
36 169 16.0 208 1 KTHY_AERPE Q9ya48 aeropyrum p
37 154 14.6 212 1 KTHY_SYNV3 Q55593 synechocyst
38 146 13.9 193 1 KTHY_THEAC Q9hlz2 thermoplasm
39 146 13.9 213 1 KTH2_SULSO P58157 sulfolobus
40 133.5 12.7 252 1 YGHR_ECOLI Q46842 escherichia
41 131.5 12.5 230 1 YGHR_ECOLI Q46844 escherichia
42 125.5 11.9 210 1 KTHY_SCHPO P36590 schizosacch
43 123.5 11.7 227 1 KTHY_MOUSE P97930 mus musculu
44 121.5 11.5 212 1 KTHY_HUMAN P23919 homo sapien
45 115.5 11.0 218 1 KTHY_CAEL Q22018 caenorhabdi

ALIGNMENTS

RESULT 1
KTHY_LACLA
ID KTHY_LACLA STANDARD; PRT; 211 AA.
AC Q9CIG4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus;
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IL1403;
RC MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006276; AAK04496.1; -
CC InterPro; IPR000062; Thymidylate_kin.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF02223; Thymidylate_kin; 1.
CC PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
FT Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 211 AA; 23999 MW; 1CC1F2C5ED9B2A55 CRC64;

Query Match 62.6%; Score 659; DB 1; Length 211;
Best Local Similarity 61.7%; Pred. No. 2.6e-40;
Matches 124; Conservative 39; Mismatches 38; Indels 0; Gaps 0;

OY 4 GFLVSLGPEGAGKTSVLEALLPILKEGVEVLITREPGGVLGKIREVILDPSTQMD 63

Db 3 GILISLEGPDGAGKTTVLKEILPEIQKMKREIVTPREPGGVRVAEIRQILDPKNTDID 62

OY 64 AKTELLLYIASRRQHLVKEKVLPALEAGKLVIMDRFIDSSVAYQGFGRLDIEADWLNOF 123

Db 63 SKTELMFAAARLHMOERMLPALQAGKVIVDFRIDSSVAYQGYGRDLGVEVWDWLNFT 122

QY 124 ATDGLKPDLTLYFDIEVEGLARIANSREVNRLDGLDHLKKVROQGYLSLLDKGNR 183

Db 123 ATDGLKPDLTLYFDVTDVALERIMKNRADEVNRLDRAEMHRKVRGEGYLEIVVKEPER 182

QY 184 IVKIDASLPLEQVETTKAVL 204

Db 183 FVKIDASQPLEKVVADTLISVL 203

RESULT 2

KTHY_BACHD

ID KTHY_BACHD STANDARD; PRT; 210 AA.

AC Q9KGL9;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE THYMIDYLATE KINASE (EC 2.7.4.9).

GN TMK OR BH0042.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C-125 / JCM 9153;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331(2000).

CC -!- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE

CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =

CC ADP + THYMIDINE 5'-DIPHOSPHATE.

CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AP001507; BAB03761.1; ..

DR InterPro; IPR000062; Thymidylate_kin.

DR Pfam; PF02223; Thymidylate_kin; 1.

DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.

KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;

KW Complete proteome.

FT NP_BIND 11 18 ATP (POTENTIAL).

FT SEQUENCE 210 AA; 23651 MW; 8F2E406148850D34 CRC64;

Query Match 52.3%; Score 551; DB 1; Length 210;

Best Local Similarity 54.8%; Pred. No. 1.2e-32;

Matches 108; Conservative 38; Mismatches 51; Indels 0; Gaps 0;

QY 1 MSKGLVLESGEGAGKTSVLEALLPILKEKGVEVLTTRPEGGVLIGKIREVILDPSPHT 60

Db 1 MTGKCFITVEGEGAGKTSALDAIEMLRENGLSVVRTRPEGGPIAQIIRSIILVDVHT 60

QY 61 QMDAKTELLLYIARRQHLVEKVLPALEAGKLVIMDRIDSSVAYQGYGRGLDIEADWL 120

Db 61 RMDPRTEALLYAARRQHLVEKVLPALEAGKLVIMDRIDSSVAYQGYGRGLDIEADWL 120

QY 121 NOFATDGLKPDLTLYFDIEVEGLARIANSREVNRLDGLDHLKKVROQGYLSLLDK 180

Db 123 ATDGLKPDLTLYFDIEVEGLARIANSREVNRLDGLDHLKKVROQGYLSLLDK 180

Db 121 NEFAIEGRYPDLTLLFRVDPDVGLSRIHRDQSQEQNRDLDOEALTFHOKVKEGYRIVET 180

QY 181 GNRIWKIDASLPLEQV 197

Db 181 PERVVEIDANQSFQV 197

RESULT 3

KTHY_BACSU

ID KTHY_BACSU STANDARD; PRT; 212 AA.

AC P37537;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).

GN TMK.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=168;

RX MEDLINE=96051385; PubMed=7584024;

RA Ogasawara N., Nakai S., Yoshikawa H.;

RT "Systematic sequencing of the 180 kilobase region of the Bacillus

RT subtilis chromosome containing the replication origin.";

RL DNA Res. 1:1-14(1994).

RN [2]

RP FUNCTION.

RA Gilles A.M., Barzu O.;

RL Unpublished observations (JUN-1996).

CC -!- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE

CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS.

CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-MONOPHOSPHATE = ADP +

CC THYMIDINE 5'-DIPHOSPHATE.

CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; D26185; BAA05264.1; ..

DR EMBL; Z99104; CAB11804.1; ..

DR Subtilist; BG10092; tmk.

DR InterPro; IPR000062; Thymidylate_kin.

DR Pfam; PF02223; Thymidylate_kin; 1.

DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.

KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;

KW Complete proteome.

FT NP_BIND 10 17 ATP (POTENTIAL).

FT SEQUENCE 212 AA; 23876 MW; 0E4519621128FA16 CRC64;

Query Match 50.1%; Score 528; DB 1; Length 212;

Best Local Similarity 51.7%; Pred. No. 5.2e-31;

Matches 106; Conservative 37; Mismatches 62; Indels 0; Gaps 0;

QY 4 GFLVLESGEGAGKTSVLEALLPILKEKGVEVLTTRPEGGVLIGKIREVILDPSPHTQMD 63

Db 3 GLFITFEGEGAGKTTVLQEKNIITAGSLQVMTATREFGGIDIAQIREVILNENILMD 62

QY 64 AKTELLLYIARRQHLVEKVLPALEAGKLVIMDRIDSSVAYQGYGRGLDIEADWLNOF 123

Db 63 PKTEALLYAARRQHLVEKVLPALEAGKLVIMDRIDSSVAYQGYGRGLDIEADWLNOF 122

QY 124 ATDGLKPDLTLYFDIEVEGLARIANSREVNRLDGLDHLKKVROQGYLSLLDKGNR 183

Db 123 AIGDMPHVTYVFSIDPEGLKRIYANGSRKRNRLDLEKDFHTKVGQYQELMKRFFER 182


```

ID KTHY_DEIRA STANDARD; PRT; 206 AA.
AC Q9RY40;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK OR DR011.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001873; AAF09698.1; -
DR HSSP; P37345; 5TMP.
DR TIGR; DR0111; -
DR InterPro; IPR000062; Thymidylate.kin.
DR Pfam; PF02223; Thymidylate.kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 11 18 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 22430 MW; E506BDFDC9F63D CRC64;

Query Match 36.8%; Score 387.5; DB 1; Length 206;
Best Local Similarity 43.4%; Pred. No. 4.7e-21;
Matches 85; Conservative 36; Mismatches 72; Indels 3; Gaps 2;

QY 1 MSKGLFVSLGPEGAGKTSVLEALLPILKEKGVEVLTTRPGGVVLIGKIREVILDPST 60
DB 1 MSQGLFITLEGPEGAGKTTQLARLEARLAAGHAVTVTRPGGTPLGTRVREVLDP-A-V 59
QY 61 OMDAKTELLYTSRRQHLVEKVLPALEAGKLVIMDRFTDSSVAYQGFGRGLDIEADWL 120
DB 60 EIEPGEFLKYSASRAQLVREVLPALENGETVLCDRVADSSAYQAGAGRLSLPLLRQI 119
QY 121 NOFATDGLKPDLTLYFDIEVEEGLARIANSREYNRLDLEGLDHLKKVRQGYLSLLDKE 180
DB 120 TAEVTTGLTPTGLTLLDLPALGLQRAARRG--QPRLEQADLTTHRRVRQGFLLAHAE 177
QY 181 GNRIVKIDASLPLEOV 196
DB 178 PQRFLVLDATREDEL 193

RESULT 7
KTHY_THEMA

```

```

ID KTHY_THEMA STANDARD; PRT; 197 AA.
AC Q9X0I3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN TMK OR TM1099.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 393:323-329(1999).
CC -!- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001769; AAD36175.1; -
DR TIGR; TM1099; -
DR InterPro; IPR000062; Thymidylate.kin.
DR Pfam; PF02223; Thymidylate.kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 7 14 ATP (POTENTIAL).
SQ SEQUENCE 197 AA; 22849 MW; 302D2EF3BAB6658E CRC64;

Query Match 35.0%; Score 368.5; DB 1; Length 197;
Best Local Similarity 42.6%; Pred. No. 9.9e-20;
Matches 81; Conservative 39; Mismatches 59; Indels 11; Gaps 4;

QY 7 VSLGPEGAGKTSVLEALLPILKEKGVEVLTTRPGGVVLIGKIREVILDPSTOMDAKT 66
DB 3 ITFEGIDSGKSTQQLLAQVLEKRGKVKILKREGPGTETGEKIRKILLE---EETPKA 59
QY 67 ELLYIAGRRQHLVEKVLPALEAGKLVIMDRFTDSSVAYQGFGRGLDIEADWLNOFATD 126
DB 60 ELFLFLASRLN-LVTEIKQYVSEGVAVLLDRYDSSVAYQGFGRNLGHEIVEELNDFATD 118
QY 127 GLKPDLTLYFDIEVEEGLARIANSREYNRLDLEGLDHLKKVRQGYLSLLDKEGNRIVK 186
DB 119 GLIPDLTYIDVDVETALKR-----KGELNR--FEKREFLERVREGVLVLAHEHPERIV 171
QY 187 IDASLPLEOV 196
DB 172 LDGKRISIEEI 181

RESULT 8
KTHY_CHLMU
ID KTHY_CHLMU STANDARD; PRT; 203 AA.
AC Q9PKK5;

```

THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
 TWK OR CT188.
 GN Chlamydia trachomatis.
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC NCBI_Taxid=813;
 RX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RA MEDLINE=J9000809; PubMed=9784136;
 RD Stephens R-S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RE Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RL Chlamydia trachomatis.";
 RS Science 282:754-759(1998).
 CC -I- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTPP IN BOTH DE
 NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
 ADP + THYMIDINE 5'-DIPHOSPHATE.
 CC -I- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 DR EMBL; AE001292; AAC67780.1; .
 DR InterPro; IPR000062; Thymidylate_kin.
 DR Pfam; PF02223; Thymidylate_kin; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
 KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 7 14 ATP (POTENTIAL).
 ST SEQUENCE 203 AA; 22473 MW; 3DACCC3690112B7C CRC64;

	Query Match	34.3%; Score 361; DB 1; Length 203;
	Best Local Similarity	40.9%; Pred. No. 3.5e-19;
	Matches	83; Conservative 40; Mismatches 60; Indels 20; Gaps 6;

```

QY      5 FLVSLEGPEGAGKTSLVEALLPILEEKGVVELTTREPGGVLIGIKIREVIDLPSTOMDA 64
       | :| |:| ||||| |:| | |:| |:| ||||| |:| |:| |:| |:| 
DB      2 FIV-VEGEGAGKTFQTALSKRLIEGRELVITTRFEGCGSLGDSVRGLLLDPQ-KISP 59
       |::|::|LAARAQHIEKIIPALKSGKTVISDRFDHFTVVVGIAAGLG-----ESFV 112
QY     65 KTELLLYASRRHLVKVPALPAEAGKLVIIMDRFIDSVAYGFGRGLDIEADMLNQFA 124
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB     60 YAEELLFLAARAQHIQEIKIIPALKSGKTVISDRFDHFTVVVGIAAGLG-----ESFV 112
QY    125 TD-----GLK---PDLTFFDIEVEGLARIANSREVNRLDEGLDLHKVKRQGYLS 175
       |: |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   113 TNLCYHVVGDKPFPLDPTIFDLLPARGLLRKA--RKHLDFKEQKPQIFHRSVREGFLA 170
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY    176 LLDKENGRIVKIDASLPLEQQWE 198
       |: |:| |:| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   171 LAEKAPRYKVIDLALLETSASVD 193
       |: |:| |:| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 10
 ID KTHY_NEIMA
 AC QJUVET; STANDARD; PRT; 206 AA.
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
 GN TWK OR NNA0869
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OC NCBI_Taxid=65699;
 RN [1]

Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecio A., Parksey D.S., Blair E., Clifton H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Masigoni V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.; "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58"; Science 287:1809-1815(2000).

RL -1- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTDP IN BOTH DE CC NOVO AND SALVAGE PATHWAYS OF DTTP SYNTHESIS (BY SIMILARITY). CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE = CC ADP + THYMINEDINE 5'-DIPHOSPHATE.

CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outpostion CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE002421; AAFA1088.1; -. DR TIGR; NMB0670; -. DR InterPro; IPR000062; Thymidylate_kin. DR Pfam; PF02233; Thymidylate_kin; I. DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE_NEG. DR Transferrase; Kinase; Nucleotide biosynthesis; ATP-binding; KW Complete proteome. FT NP_BIND 10 17 ATP (POTENTIAL). SQ SEQUENCE 206 AA; 22986 MW; 2D356EFE373E8F5E CRC64;

Query Match 33.8%; Score 356; DB 1; Length 206;
Best Local Similarity 40.8%; Pred. No. 8e-19;
Matches 82; Conservative 38; Mismatches 77; Indels 4; Gaps 3;

QY 7 VSLEGPCKAKTSVLKALLPIEEKGVELVTREPGGVLIGEIKREVLDPSHTOMDAKT 66
::|:||||: |: :||: ||||||| ::|||:| |::|:| |:
Db 6 ITLDGDGAGKSTNLAVIKAWFERRGFLVLTREPGGTVPEALREILLNP-ETKAGLA 64

QY 67 ELLLYIASRRHQHVEKLVPALKGLKLVIMDRITSSVAYOGFGRLDIEAIDLNLPATD 126
| |:| |:| :|:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 65 ETLMFAARMQHLEEVILPALSDGIHVVSORFDATFAYOGGGRGMSPSEDIILEHWQG 124

QY 127 GLKPDLTYLFIDEVEEGLARIAANSDEVNRNLDEGLDHKKVRQGYLSLLDKENRIVK 186
||||||| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 125 GLKPDLTLLEDVPLEYSMRI--GTQEKDFEQEQANDFFMRVGVYLDRAACPERYAV 182

QY 187 IDASLPLEQQVETTKAVLFDG 207
|::: |:|:| |:|:| |::|
Db 183 IDSNNRLDEVNRNSTEKL-DG 202

RESULT 12
KTHY_CAUCR ID KTHY_CAUCR STANDARD; PRT; 208 AA.
AC Q9OJ9; .
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE). TMK OR OC1824.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CH15:


```
RA Ohta N., Newton A.;
RT "dnaC encodes the delta prime subunit of DNA polymerase III."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CBL5;
RC MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.K.R., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Esmolaeva M., White O.,
RA Salzberg S.L., Shapiro L., Venter J.C., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF099189; AAF06831.1; -
DR EMBL: AF005856; AAK23799.1; -
DR TIGR: CC1824; -
DR InterPro: IPR00062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1
DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 11 18 ATP (POTENTIAL).
SQ SEQUENCE 208 AA; 22107 MW; 7BD15DFD05F42D02 CRC64;

Query Match 32.08; Score 337; DB 1; Length 208;
Best Local Similarity 37.68; Pred. No. 1.8e-17;
Matches 79; Conservative 34; Mismatches 95; Indels 2; Gaps 1;

QY 1 MSKGLVSLGEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLGKIREVILDPST 60
DB 1 MTQGFIFTEGEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLGKIREVILDPST 60

QY 61 QMDAKTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDTEAIDWL 120
DB 61 RWSPTESLLMYAARRDHIERVIRGLARGAVVLCDFREADSTRAYQAGGDAPASLTAAL 120

QY 121 NQFATDGLKPDLTLYFDIEVEGLARIANSREVRNLDLEGLDLHKVQGYLSLDKE 180
DB 121 EHVILGSGTVPVLTLLDPAEVLQRAARG--AARFESKGLAFHERLRAGYLEIARRE 178

QY 181 GNRIVKIDASLPLEQVQVETTKAVLFDGML 210
DB 179 PDRCVIDAADAEALDAVTAISDVVVQRLGL 208

RESULT 13
KTHY_CHLPN
ID KTHY_CHLPN STANDARD; PRT; 206 AA.
AC Q928R5; Q9J094;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (DTPP KINASE).
GN TWK OR CPN0273 OR CP0486.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
```

```
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CWL029;
RC MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RN Nat. Genet. 21:385-389(1999).
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=J138;
RC MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001612; AAD18422.1; -
DR EMBL: AE002210; AAF38316.1; -
DR EMBL: AP002346; BAA98483.1; -
DR TIGR: CP0486; -
DR InterPro: IPR00062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1
DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 7 14 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 22852 MW; 42B0E8AA5262768 CRC64;

Query Match 31.38; Score 329.5; DB 1; Length 206;
Best Local Similarity 35.38; Pred. No. 6.1e-17;
Matches 76; Conservative 46; Mismatches 72; Indels 21; Gaps 5;

QY 5 FLVSLGEGAGKTSVLEALLPILKEKGVLEVTTRPGGVLGKIREVILDPSTOMDA 64
DB 2 FIV-IEGEGSGKSLAKALGDQLVADQKRVLLTRPGGCGGLIGLRDLILPPHLELSR 60

QY 65 KTELLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDTEAIDWLNOFA 124
DB 61 CCELFELGSRQHQIQLPVRDGYIVICERFHSITIVQGIAGLAD-----FV 113

QY 125 TD-----GLKPDLTLYFDIEVEGLARIANSREVRNLDLEGLDLHKVQGYLS 175
DB 114 ADLCSKVVGPTTFFLPNLFVLLDIPADIGLQR--KHKQKVFQKFKPLSYHNIREGFLS 171
```

```
Qy 176 LLDKFGNRIVKIDASLPLEQWVETTKAVLDFCMGL 210
      | : | : | | | : : | : | : | |
Db 172 LASADPSRYLVLDARESLASLID--KVMLHTQLGL 204
```

RESULT 14		
KTHY_PASMU		
ID	KTHY_PASMU	STANDARD; PRT; 209 AA.
AC	Q3CKE9;	
DT	20-AUG-2001	(Rel. 40, Created)
DT	20-AUG-2001	(Rel. 40, Last sequence update)
DT	20-AUG-2001	(Rel. 40, Last annotation update)
DE	THYMIDYLATE KINASE (EC 2.7.4.9)	(DTMP KINASE
GN	TNK OR Pw1673.	

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;

RP SEQUENCE FROM N.A.

RX MEDLINE=21145866: PubMed=11248100:

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of *Pasteurella multocida* Pm70.";

Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -I- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM DTPD IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).

CC -I- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =
CC ADP + THYMIDINE 5'-DIPHOSPHATE.

CC -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; AE006204; AAK03757.1; ALT INIT.

DR InterPro; IPR000062; Thymidylate_kin.

DR Pfam; PF02223; Thymidylate_kin; 1.

DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.

KW **Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;**

KW Complete proteome.

FT	NP_BIND	11	18	ATP (POTENTIAL).
----	---------	----	----	------------------

SQ SEQUENCE 209 AA; 23480 MW; 96CFF095398A8FD2 CRC64;

Query Match

Best Local Similarity 39.38; Pred. No. 7.8e-16;

Matches 81; Conservative 38; Mismatches 81; Indels

QY 1 MSKGLFVLSLEGPGAGKTSVLEALLPILEEKGV-EVLITREPGGVLIGKIREVILDPH 59
| : | ||||| : :: | : :: ||||| : ||::| :
Dd 1 MTTGKFIVLEGEGAGKTARDSIVRALHANGIHDIYETREPGGTPLAKRLRIKHETE 60

60 TQMDAKTELLVYASRRQHLEKVL-PALEAGKLVMDFIDSSVAYQGFRGLDIEAID 118
: : : : : : : : : : : : : : : : ;
61 EPVTDKAEMLLYAARIO-LVENVIKPALAQGKKWTGDRHDMSSOAYOGGGRLDOHLH 119

QY	119	WLNQFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDEGLDLHKVKVRCQYLSLLD	178
		: : : : :: : : : :	
Db	120	TLKQTILGEFEPTLYLIDIPVLGSR--AKRGALDRIEQNQLDFHTRQRQVELV-	176

QY 179 KEGNRIVKIDASLPLEQVWETTKAVL 204

[illegible]

Dbb 177 RHNPKAVTIDASQTMSKVAEDVESAI 202

RESULT 15
KTHY_VIBCH

1
2
3
4
5

ID	KTHY_VIBCH	STANDARD;	PRT;	212 AA.
AC	Q9KQ12;			
DR	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DI	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	THYMIDYLATE KINASE (EC 2.7.4.9) (DMP KINASE).			
DN	TMK OR VC2016.			
GN	Vibrio cholerae.			
OS	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OC	NCBI_TaxID=666;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	RC STRAIN=EL TOR NI6961 / SEROTYPE O1;			
RF	MEDLINE=20406833; PubMed=10952301;			
RX	Heidelberg J.F., Eelsen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RB	Gill S.R., Nelson K.E., Read T.D., Tetzelin H., Richardson D.,			
RC	Emoleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RD	McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,			
RE	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RF	Fraser C.M.;			
RG	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RI	cholerae";			
RL	Nature 406:477-483(2000).			
RM	-1- FUNCTION: PHOSPHORYLATION OF DTMP TO FORM dTDP IN BOTH DE			
RN	NOVO AND SALVAGE PATHWAYS OF dTTP SYNTHESIS (BY SIMILARITY).			
RO	-1- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-PHOSPHATE =			
RP	ADP + THYMIDINE 5'-DIPHOSPHATE.			
RQ	-1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.			
RS	-----			
RT	This SWISS-PROT entry is copyright. It is produced through a collaboration			
RU	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
RV	the European Bioinformatics Institute. There are no restrictions on its			
RW	use by non-profit institutions as long as its content is in no way			
RX	modified and this statement is not removed. Usage by and for commercial			
RY	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
RZ	or send an email to license@isb-sib.ch).			
SA	-----			
SB	EMBL; AE004276; AAF95164.1; .			
SC	DR TIGR; VC2016; .			
SD	InterPro; IPR000062; Thymidylate_kin.			
SE	Pfam; PF02223; Thymidylate_kin; 1.			
SF	DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.			
SG	KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;			
SH	KW Complete proteome.			
SI	FT NP_BIND 10 17 ATP (POTENTIAL).			
SJ	SQ SEQUENCE 212 AA; 23666 MW; 8f4F51A00CE60F4D CRC64;			
SK				
SL	Query Match 29.8%; Score 314; DB 1; Length 212;			
SM	Best Local Similarity. 39.5%; Pred.No.7.9e-16;			
SN	Matches 83; Conservative 42; Mismatches 75; Indels 10; Gaps 8;			
SO				
SP	QY 1 MSKGFLVLEGPAGAKTSVLEALLPLIEKGVLEVLT-TREPQGLIGEKIREVILD--P 57			
SR	I :			
SS	I :			
ST	Db 1 MNKFIV-IESLEGAGKSTAIQVVVFTELOQNGDHITRTREPQGTLLAEKLALVKKEHP 59			
SV	I :			
SW	QY 58 SHQMOKATTELLLYIASRRHLLVEKVU-PALEAGKLVIMDRFDISVAVQFGGRLDIEA 116			
TX	I :			
SY	Db 60 GEELQDT-TELLLYVAARVQ-LVENVIKPALARGEWVGVDGRHDMSQAQGGGQIAPST 117			
SZ	I :			
TA	QY 117 IDMLNQFATDGKLPDLTFYDFIEVEEGLARIANSDEVNRLDLEGLDLHKVRQGVLSL 176			
TB	I :			
TC	Db 118 MQSLKQTALGDFFPKLFYLDIDIPKGLER--ARGGELDRIENKMDSIFFERARYLEL 175			
TD	I :			
TE	QY 177 LDREGNRIVKIDASLPLEQVVTTKAVLFD 206			
TF	I :			
TG	Db 176 ANSD-DSVMVIDAAQSIEQVTADIRRALQD 204			
TH	I :			

Search completed: February 15, 2002, 02:02:25
Job time: 388 sec

Search completed: February 15 2002 02:02:25

Job time: 388 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 01:55:17 ; Search time 55.75 Seconds
(without alignments)
556.228 Million cell updates/sec

Title: US-09-749-972-2
Perfect score: 1053
Sequence: 1 MSKGFVLSLEGAGKTSV.....LEQVVETTKAVLFDGMGLAK 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTEMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_prodent.*
12: sp_virus.*
13: sp_invertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670.5	63.7	211	Q9A190	Q9A190 streptococ
2	653	62.0	211	Q9A0D5	Q9A0D5 lactococ
3	475	45.1	205	Q9WCL1	Q9WCL1 staphyloc
4	373	35.4	1089	Q9X908	Q9X908 streptomyce
5	365	34.7	226	Q9EZH7	Q9EZH7 xanthomonas
6	233	22.1	165	Q9R2M7	Q9R2M7 versinia en
7	232	22.0	165	Q9R2M9	Q9R2M9 versinia en
8	232	22.0	165	Q9R2M8	Q9R2M8 versinia en
9	231	21.9	165	Q9R1B4	Q9R1B4 versinia ps
10	230	21.8	165	Q9R1E6	Q9R1E6 versinia ps
11	230	21.8	165	Q9R2Y0	Q9R2Y0 versinia ps
12	201	19.1	98	Q49001	Q49001 mycoplasma
13	174	16.5	208	Q83373	Q83373 treponema p
14	145	13.8	257	Q51733	Q51733 borrelia bu
15	144	13.7	309	Q9C2R7	Q9C2R7 neurospora
16	141	13.4	388	12 Q9DQD0	Q9DQD0 white spot
17	134.5	12.8	212	11 Q9D7I3	Q9D7I3 mus musculus
18	129	12.3	188	10 Q81650	Q81650 arabidopsis
19	120.5	11.4	212	4 Q9BUX4	Q9BUX4 homo sapien

20	107	10.2	80	2	Q9A29	Q9A29 vibrio mari
21	106.5	10.1	210	2	Q9CCJ3	Q9CCJ3 mycobacteri
22	103.5	9.8	205	2	Q9W52	Q9W52 staphylococ
23	103	9.8	333	4	Q9Y804	Q9Y804 aeropyrum p
24	103	9.8	903	4	Q9BWL4	Q9BWL4 homo sapien
25	102	9.7	310	2	Q9PE05	Q9PE05 xylella fas
26	101	9.6	168	3	Q9HFB8	Q9HFB8 thanatephor
27	101	9.6	337	1	Q9YDH3	Q9YDH3 aeropyrum p
28	101	9.6	405	1	Q28871	Q28871 archaeoglob
29	100	9.5	248	1	Q9YCF6	Q9YCF6 aeropyrum p
30	100	9.5	314	2	Q9KEN4	Q9KEN4 bacillus ha
31	99.5	9.4	250	1	Q58948	Q58948 pyrococcus
32	99.5	9.4	288	2	Q9KF34	Q9KF34 bacillus ha
33	99.5	9.4	328	1	Q9V0X0	Q9V0X0 pyrococcus
34	98.5	9.4	1482	3	Q74637	Q74637 penicillium
35	98	9.3	239	2	Q31427	Q31427 bacillus su
36	98	9.3	608	1	Q9UZN6	Q9UZN6 pyrococcus
37	97.5	9.3	214	2	Q05891	Q05891 mycobacteri
38	97.5	9.3	1499	3	Q9Y839	Q9Y839 mycosphaere
39	97	9.2	265	2	Q9X665	Q9X665 staphylococ
40	97	9.2	605	1	Q58850	Q58850 pyrococcus
41	96.5	9.2	171	5	Q9V8B5	Q9V8B5 drosophila
42	96.5	9.2	558	2	Q9JR75	Q9JR75 neisseria m
43	95.5	9.1	503	2	Q9CLB3	Q9CLB3 pasteurella
44	95.5	9.1	1619	3	Q13407	Q13407 magnaporthe
45	95	9.0	374	1	Q46492	Q46492 acidianus a

ALIGNMENTS

RESULT 1
Q9A190 PRELIMINARY; PRT; 211 AA.
ID AC Q9A190;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE THYMIDYLATE KINASE.
GN TMK OR SPY0399.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvarov A.N., Kenton S., Lal H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: A9006502; AAK33434.1;
KW Kinase; Complete proteome.
SQ SEQUENCE 211 AA; 23402 MW; 3866FBI39A188752 CRC64;

Query Match	63.7%	Score	670.5;	DB	2;	Length	211;
Best Local Similarity	67.7%	Pred. No.	4.8e-39;				
Matches	134;	Conservative	30;	Mismatches	33;	Indels	1;
Gaps	1;						
QY	1	MSKGFVLSLEGAGKTSVLEALLPILEK-GVEVLTTREPGGVLGEKIREVILDP	59				
Db	1	MITGKLITVGGPDGAGTTLVLEQLIPLLKQKVAQDILTTREPGGVAISEHREIL	60				
QY	60	TQMDAKTELLYIASRRQHLVEKVLPALEAGKLVIMDRFDISSVAYQGFGLDIEA	119				
Db	61	TAMPDKTELLYIAARRQHLVEKVLPALEAGQLVIMDRFDISSVAYQGFGLKADIQ	120				
QY	120	LNQFATDGLKPDLTLYFDIEVEEGLARIANSREVNRLDLEGLDHLKKVQGLSL	179				

Db 121 LNEFATDGLPDLTLYFDVPSEIGLARINANOQREVNRLDLETIEHQRVKGYLALAKE 180
QY 180 EGNRVKIDASLPLEQV 197
Db 181 HPKRIVTIDATKPLEV 198

RESULT 2
Q9AQD5 ID Q9AQD5 PRELIMINARY; PRT; 211 AA.
AC Q9AQD5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TSCORF1
GN TSCORF1984.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGI363;
RX MEDLINE=20244638; PubMed=10784052;
RA O'Connell-Motherway M., van Sinderen D., Morel-Deville F.,
Fitzgerald G.F., Ehrlich S.D., Morel P.;
RT "Six putative two-component regulatory systems isolated from
Lactococcus lactis subsp. cremoris MGI363.";
RL Microbiology 146:935-947(2000).
DR EMBL; AF178425; AAG53731.1; -.
SQ SEQUENCE 211 AA; 23973 MW; 9AADB5F3F6985EFA CRC64;

Query Match 62.0%; Score 653; DB 2; Length 211;
Best Local Similarity 62.2%; Pred. No. 7.8e-38;
Matches 125; Conservative 36; Mismatches 40; Indels 0; Gaps 0;
QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVVLTTRPFGVGLIGKEKREVILDPSSHOMD 63
Db 3 GILSLGPDGAGTTLVQLPEIQMKREVTPTRPFGGVVAEIRQIILDPKNTIED 62
QY 64 AKTELLYIARRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADLWNOF 123
Db 63 SKTEMLFAAARLHMQLPALRACKVIVDFIDSSVAYQYGRDLGVVVDLWNYF 122
QY 124 ATDLKLPDLTLYFDIEVEGLARIAANSRDREVNRLDLEGLDHLKKVQGYLSLLDKGNR 183
Db 123 ATDLKLPDLTLYFDITDVALERTMKNRADEVNRLDLERAEMHRKVRGYLEIVAKPEGR 182
QY 184 IVKIDASLPLEQVETTKAVL 204
Db 183 FVKIDASQSLKVVADTLEVL 203

RESULT 3
Q99WC1 ID Q99WC1 PRELIMINARY; PRT; 205 AA.
AC Q99WC1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TWK PROTEIN
GN TWK OR SA0440.
OS Staphylococcus aureus subsp. aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,

RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003130; BAB41670.1; -.
KW Complete proteome.
SQ SEQUENCE 205 AA; 23424 MW; 71F9C96511FF3A5E CRC64;

Query Match 45.1%; Score 475; DB 2; Length 205;
Best Local Similarity 49.5%; Pred. No. 1.4e-25;
Matches 96; Conservative 40; Mismatches 54; Indels 4; Gaps 3;
QY 7 VSLGPEGAGKTSVLEALLPILEKGVVLTTRPFGVGLIGKEKREVILDPSSHOMD 66
Db 5 ITFEGPEGSGKTVTVNEVYHRL-VKDYDVTMTREPGGVPTGEIRKIVLEGN--DMDIRT 61
QY 67 ELLYIARRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGFGRGLDIEADLWNOF 126
Db 62 EAMLFAASRRHRLVLPALKEGKVLCDRYIDSSLAYQGVARGIGVEEVRLNEFAIN 121
QY 127 GLKLPDLTLYFDIEVEGLARIAANSRDREVNRLDLEGLDHLKKVQGYLSLLDKGNR 186
Db 122 GLYDPTLYLVNVAEVRGRIKNS-RDQNRDQDLKFKHEKVIQEGYQEIHHNESQRPKS 180
QY 187 IDASLPLEQVETTKAVL 200
Db 181 VNAQPLENVVETD 194

RESULT 4
Q9X908 ID Q9X908 PRELIMINARY; PRT; 1089 AA.
AC Q9X908;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE INTEGRAL MEMBRANE PROTEIN WITH KINASE ACTIVITY.
GN SCH5.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL035636; CAB38479.1; -.
DR HSSP; P37345; STMP.

DR InterPro; IPR003662; sub.transporter.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.

```
KW Kinase.
SQ SEQUENCE 1089 AA; 115739 MW; F75DD5B2055F409E CRC64;

Query Match 35.4%; Score 373; DB 2; Length 1089;
Best Local Similarity 40.9%; Pred. No. 1.5e-17;
Matches 81; Conservative 35; Mismatches 78; Indels 4; Gaps 1;

QY 4 GFLVSLGPEGAGKTSVLEALLPILEKGVVLTTRPGGVLTGKIREVILDPSTQMD 63
   ||:||||:||||: ||| : || ||: ||||| :|||: ||| |
Db 509 GFFIALEGGDAGKSTQAEALAEWIRGKGHEWLTREPGATPVGKRLRSILLDVSSAGLS 568

QY 64 AKTELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQFGRLDTEADWLNQF 123
   ||:||||: ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 569 HRAEALLYADRAEHVDTVWRPALERGAVVSDRYIDSSVAYQAGRDLSPTETARINRW 628

QY 124 ATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDLEGLDHLKHKVGYLSLLDKEGNR 183
   ||:||||: ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 629 ATNGLVHLTVLLDVAPAEARETEREAPD---RLESEPAEFHARVRSGLTFLAANDPGR 684

QY 184 IVKIDASLPLEQVYETTK 201
   : || | | :
Db 685 YLVVDAGQEPVAVTVVR 702

RESULT 5
QYEQZH7 PRELIMINARY; PRT; 226 AA.
AC QYEQZH7;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE THYMIDYLATE KINASE.
GN TMK.
OS Xanthomonas albilineans.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=29447;
RN SEQUENCE FROM N.A.
RP MEDLINE=20547559; PubMed=11094291;
RX Huang G., Zhang L., Birch R.G.;
RA "Characterization of the acyl carrier protein gene and the fab gene
RT locus in Xanthomonas albilineans.";
RL FEMS Microbiol. Lett. 193:129-136(2000).
DR EMBL; AF294440; A642374.1; -.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
KW Kinase.
SQ SEQUENCE 226 AA; 24882 MW; F88D6E277DB7BECF CRC64;

Query Match 34.78%; Score 365; DB 2; Length 226;
Best Local Similarity 43.1%; Pred. No. 6.3e-18;
Matches 93; Conservative 35; Mismatches 74; Indels 14; Gaps 5;

QY 1 MSKGFL-----VSLEGPEGAGKTSVLEALLPILEKGVVLTTRPGGVLTGKIREVIL 55
   ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MSEAVLRHRRVSLGEGGAGKTAIRNDRWLQAQHEVLTTRPGGTPAERIRGLLL 60

QY 56 D--PSHTQ-----MDAKTELLLYIASRRQHLVEKVLPALEAGKLVIMDRFIDSSVAYQGF 108
   ||: ||| : ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 DMAPSSPETAVERLSAETELLVFAARAHVREVIRPALQRGAVVSDRFTDSSVAYQGE 120

QY 109 GRGLDIEADWLNQFATDGLKPDLTLYFDEVEEGLARIAANSREVNRLDLEGLDHLK 168
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 GRGLDRAWIADLERRAV--GLQPGTLLDLDVQIGRAR--TSGRDLWPDRIESEQDDFFQR 178

QY 169 VRGYLSLLDKEGNRIVKIDASLPLEQVWETTKAVL 204
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 VRAGFORATQDPQFRFRVIDASQPQVAQAANAAL 214
```

```
RESULT 6
QYR2M7 PRELIMINARY; PRT; 165 AA.
AC QYR2M7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN TMK.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN SEQUENCE FROM N.A.
RP STRAIN=21506, 25963, AND 21708;
RX MEDLINE=20040673; PubMed=10570195;
RA Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;
RT "Yersinia pestis, the cause of plague, is a recently emerged clone of
RT pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL; AJ270425; CAB58190.1; -.
DR EMBL; AJ270423; CAB58188.1; -.
DR EMBL; AJ270424; CAB58189.1; -.
DR HSSP; P37345; STMP.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Kinase; Transferase.
FT NON_TER 1
FT NON_TER 165
SQ SEQUENCE 165 AA; 18220 MW; F9BBEC3F61F2C8EB CRC64;
```

```
Query Match 22.1%; Score 233; DB 2; Length 165;
Best Local Similarity 38.6%; Pred. No. 5.2e-09;
Matches 61; Conservative 30; Mismatches 61; Indels 6; Gaps 5;

QY 22 EALLPILEKGV-EVLTTRPGGVLTGKIREVILDPSTQ-MDAKTELLLYIASRRQHL 79
   : : : | : : : : ||||| : ||| : : : | : : | : | | |
Db 3 DTVVATLRAQGINDIVTREPGGTPAEKLDKDLKQGDGDEVLTDRKAEVLMLYAARVQ-L 61

QY 80 KEVL-PALEAGKLVIMDRFIDSSVAYQFGRLDIEADWLNQFATDGLKPDLTLYFDI 138
   ||: ||| | | | | | | | | | | | | | | | | | | | | |
Db 62 VENVIKPALARGSWVGDRLDSSQAYOGGGRGIDINLMTSLRDIVLQGFEPDLTLYLDL 121

QY 139 EVEGLARIAANSREVNRLDLEGLDHLKHKVQGYLSL 176
   |||| | ||| : | : | : | : | : | : | : | : |
Db 122 PPAIGLARARG--ELDRIEQESLAFFPTRERYLEL 157
```

```
RESULT 7
QYR2M9 PRELIMINARY; PRT; 165 AA.
AC QYR2M9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN TMK.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN SEQUENCE FROM N.A.
RP STRAIN=24636, AND 383;
RX MEDLINE=20040673; PubMed=10570195;
RA Achtman M., Zurth K., Morelli G., Torrea G., Guiyoule A., Carniel E.;
RT "Yersinia pestis, the cause of plague, is a recently emerged clone of
RT pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL; AJ270428; CAB58193.1; -.
```

DR	EMBL; AJ270426; CAB58191.1; -.
DR	HSSP; P37345; STMP.
DR	InterPro; IPR000062; Thymidylate_kin.
DR	Pfam; PF02223; Thymidylate_kin; 1.
DR	PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW	Kinase; Transferase. 1
FT	NON_TER 1 165
FT	NON_TER 165 165
SQ	SEQUENCE 165 AA; 18247 MW; F9A5F8B55AF2C8EB CRC64;

Query Match	22.0%; Score 232; DB 2; Length 165;
Best Local Similarity	38.6%; Pred. No. 6.1e-09;
Matches 61; Conservative 30; Mismatches 61; Indels 6; Gaps	

QY	22 EALLPILEEGV-EVLTTREPGGVGIGEKIREVILDPSTHQ-MDAKTELLLYIASRRQHL 79
Dd	: : : : : : : : : : : : : : : : :
Dd	3 DTVVTATRAAGINDIVTREPGGTPLAEKLRLDIKOGIDGEVLTKAEVLMLYAARVQ-L 61
QY	80 KEVKL-PALAEAGKLIVIMDRFDISSVAYQGFGRLDIEADWLNQFATDGKLPDLTYFYDI 138
Dd	: : : : : : : : : : : : : : :
Dd	62 VENVIKPALARGSWWGDHRLDSQAQQGGGRGIDINLMTSLRDTVLGFEFRPDLTLYLDL 121
QY	139 EVEEGLAIANSRVNRLDLEGLDLHKKVROGYLSL 176
Dd	: : : : : : : : :
Dd	122 PPVIGLARARARG--ELDRIEQEQLAFFERTTRERYLKL 157

RESULT	8
Q9R2M8	
ID	Q9R2M8 PRELIMINARY; PRT; 165 AA.
AC	Q9R2M8;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN	TWK.
OS	Yersinia enterocolitica.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX	Yersinia.
OX	NCBI_TaxID=630;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN-WA, AND YER081;
RX	MEDLINE=20040673; PubMed=10570195;
RA	Achtman M., Zurth K., Morelli G., Torrea G., Gulyoule A., Carniel E.;
RT	"Yersinia pestis, the cause of plague, is a recently emerged clone of
RT	pseudotuberculosis";
RL	Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR	EMBL; AJ270429; CAB58194.1; -.
DR	EMBL; AJ270427; CAB58192.1; -.
DR	HSSP; P37345; STMP.
DR	InterPro; IPR000062; Thymidylate_kin.
DR	Pfam; PF02223; Thymidylate_kin; 1.
DR	PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW	Kinase; Transferase.
FT	NON_TER 1 165
FT	NON_TER 165 165
SQ	SEQUENCE 165 AA; 18248 MW; F9A84EB55AF2C8EB CRC64;

Query Match	22.0%; Score 232; DB 2; Length 165;
Best Local Similarity	38.6%; Pred. No. 6.1e-09;
Matches 61; Conservative 30; Mismatches 61; Indels 6; Gaps	

QY	22 EALLPILEEGV-EVLTTREPGGVGIGEKIREVILDPSTHQ-MDAKTELLLYIASRRQHL 79
Dd	: : : : : : : : : : : : : : : : :
Dd	3 DTVVTATRAAGINDIVTREPGGTPLAEKLRLDIKOGIDGEVLTKAEVLMLYAARVQ-L 61
QY	80 KEVKL-PALAEAGKLIVIMDRFDISSVAYQGFGRLDIEADWLNQFATDGKLPDLTYFYDI 138
Dd	: : : : : : : : : : : : : : :
Dd	62 VENVIKPALARGSWWGDHRLDSQAQQGGGRGIDINLMTSLRDTVLGFEFRPDLTLYLDL 121

```
RT pseudotuberculosis";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL: AJ270417; CAB58204.1; -.
DR HSP; P37345; STMP.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1.
KW Kinase; Transferase.
FT NON_TER 1
FT NON_TER 165
SQ SEQUENCE 165 AA; 18103 MW; 6D294FB9A075EBE CRC64;

Query Match 21.88; Score 230; DB 2; Length 165;
Best Local Similarity 38.08; Pred. No. 8.4e-09;
Matches 60; Conservative 31; Mismatches 61; Indels 6; Gaps 5;

QY 22 EALLPILEEGV-VLTTREPGGVIGEKIREVILDPSTQ-MDAKTELLLYIASRQHL 79
Db 3 DTVAVLURAGINDIVFTREPGGTPLAEKURDLIKQIDGGEVLTKAEVLMYAAVQ-L 61
QY 80 VEKVL-PALEAGKLVIMDRFTDSSVAYQGFGRGLDIEAIDWLNFATDGLKPDLTLYFDI 138
Db 62 VENVIKALARGSWVGDRLDLSQAYQGGGIDSQLMASLRDTVLGFEFRPDLTLYLDL 121
QY 139 EVEGLARIAANSREVRNRLDLEGLDLHKVKVQGYLSL 176
Db 122 PPAVGLARARG--ELDRIEQESLAFFERTRARYLEL 157

QY 80 VEKVL-PALEAGKLVIMDRFTDSSVAYQGFGRGLDIEAIDWLNFATDGLKPDLTLYFDI 138
Db 62 VENVIKALARGSWVGDRLDLSQAYQGGGIDSQLMASLRDTVLGFEFRPDLTLYLDL 121
QY 139 EVEGLARIAANSREVRNRLDLEGLDLHKVKVQGYLSL 176
Db 122 PPAVGLARARG--ELDRIEQESLAFFERTRARYLEL 157

RESULT 11
Q9R2Y0
ID Q9R2Y0 PRELIMINARY; PRT; 165 AA.
AC Q9R2Y0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (EC 2.7.4.9) (FRAGMENT).
GN TMK.
OS Versinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=31830, 32937, 32953, AND 32934;
RX MEDLINE=20040673; PubMed=10570195;
RA Achtnan M., Zurth K., Morelli G., Torrea G., Guilyoule A., Carniel E.;
RT "Versinia pestis, the cause of plague, is a recently emerged clone of
pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14043-14048(1999).
DR EMBL: AJ270418; CAB58225.1; -.
DR EMBL: AJ270419; CAB58210.1; -.
DR EMBL: AJ270420; CAB58211.1; -.
DR EMBL: AJ270421; CAB58212.1; -.
DR HSP; P37345; STMP.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; 1.
KW Kinase; Transferase.
FT NON_TER 1
FT NON_TER 165
SQ SEQUENCE 165 AA; 18073 MW; CD385EB9ACD22869 CRC64;

Query Match 21.88; Score 230; DB 2; Length 165;
Best Local Similarity 38.08; Pred. No. 8.4e-09;
Matches 60; Conservative 31; Mismatches 61; Indels 6; Gaps 5;

QY 22 EALLPILEEGV-VLTTREPGGVIGEKIREVILDPSTQ-MDAKTELLLYIASRQHL 79
Db 3 DTVAVLURAGINDIVFTREPGGTPLAEKURDLIKQIDGGEVLTKAEVLMYAAVQ-L 61

QY 80 VEKVL-PALEAGKLVIMDRFTDSSVAYQGFGRGLDIEAIDWLNFATDGLKPDLTLYFDI 138
Db 62 VENVIKALARGSWVGDRLDLSQAYQGGGIDSQLMASLRDTVLGFEFRPDLTLYLDL 121
QY 139 EVEGLARIAANSREVRNRLDLEGLDLHKVKVQGYLSL 176
Db 122 PPAVGLARARG--ELDRIEQESLAFFERTRARYLEL 157

RESULT 12
Q49001
ID Q49001 PRELIMINARY; PRT; 98 AA.
AC Q49001;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ATP-BIND. PYRIMIDINE KINASE (FRAGMENT).
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343(KID);
RX MEDLINE=96059641; PubMed=7476192;
RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,
RA Gilbert W., Gillevet P.M.;
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
its physiology.";
RL Mol. Microbiol. 16:955-967(1995).
DR EMBL: Z33079; CAA83746.1; -.
DR HSP; P37345; STMP.
FT NON_TER 98
SQ SEQUENCE 98 AA; 11449 MW; FA597FA1685A0822 CRC64;

Query Match 19.18; Score 201; DB 2; Length 98;
Best Local Similarity 51.28; Pred. No. 4.1e-07;
Matches 42; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

QY 7 VSLGPEGAGTSVLEALLPILEEGVVLTTREPGGVIGEKIREVILDPSTQMDAKT 66
Db 3 ITFEGMDGSGKTTALLKVKEELRNLYKVLITREPGGEVIAEQIRIILDNKNKMDAWT 62
QY 67 ELLLYIASRQHLVEKVLPALE 88
Db 63 EALLFIASRQHLQKVIKPALE 84
QY 67 ELLLYIASRQHLVEKVLPALE 88
Db 63 EALLFIASRQHLQKVIKPALE 84

RESULT 13
O83373
ID O83373 PRELIMINARY; PRT; 208 AA.
AC O83373;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE THYMIDYLATE KINASE (TMK).
GN TP0354.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ufferback T.,
RA McDonald L., Ariach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts R., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
```

[illegible]

<hr/>					
Query Match	16.5%;	Score 174;	DB 2;	Length 208;	
Best Local Similarity	30.0%;	Pred. No. 8.4e-05;			
Matches 61; Conservative	36;	Mismatches 80;	Indels 26;	Gaps 8;	
<hr/>					
QY	7	VSLEGPAGKTSVLEALLPILKEKGVEVLTTREPGVGLIGEKIREVILDPSHTQMDAKT	66		
		: : : : :			
Dd	8	VWFEIGDGTGTSQLRALERHFOAR-KDMVFTQPTGGTEIGTLIRDVL--QKRVISSKA	64		
		: : : : :			
QY	67	ELLYLTASRRQHL--VEKVLPALEACKLVIMDRFIDSSVAYQGFGRLDEAIDLNLNQFA	124		
		: : : : : : :			
Dd	65	LGLLFRADRHEHLEGAGGINDCLAEGKIVLCRDYVFSSLYIQGM-----AVS--GSFA	115		
		: : : : : : :			
QY	125	TDLKL----PDLTLYFDIEVEEGLARIANSDEVNRLDLEGLDLHKVKRQGYLSLLDK-	179		
		: : : : : : :			
Dd	116	YELNKEFPLEVFFYADPIEVCEVRITA---RLQTLEYVTSTFQEKARKKYETIFRC	172		
		: : : : : : :			
QY	180	---EGNRIVKIDASLPLEQVVE 198			
		: : : : : : :			
Dd	173	RHLYPAMKVIEDAREEIVVHE 195			
		: : : : : : :			
<hr/>					
RESULT 14					
O51733					
ID O51733	PRELIMINARY;	PRT;	257 AA.		
AC O51733;					
AT Q1-JUN-1998 (TREMBLrel. 06, Created)					
<hr/>					
RESULT 15					
Q9C2R7					
ID Q9C2R7	PRELIMINARY;	PRT;	309 AA.		
AC Q9C2R7;					
DT 01-JUN-2001 (TREMBLrel. 17, Created)					
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE RELATED TO THYMIDYLATE KINASE.					
GN 104H10..70.					
OS Neurospora crassa.					
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;					
OC Sordariales; Sordariaceae; Neurospora.					
OX NCBI_Taxid=5141;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,					
RA Nyakatura G., Mewes H.W., Mannhaupt G.;					
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.					
RN [2]					
RP SEQUENCE FROM N.A.					
RA German Neurospora genome project;					
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.					
DR EMBL; AL513410; CAC28573.1; -.					
KW Kinase.					
SQ SEQUENCE 309 AA; 33757 MW; AC2E55FC6B795DD1 CRC64;					

[illegible]

Search completed: February 15, 2002, 02:01:43
Job time: 386 sec

Query Match 13.8%; Score 145; DB 2; Length 257;
Best Local Similarity 26.7%; Pred. NO. 0.011;
Matches 48; Conservative 35; Mismatches 79; Indels 18; Gaps 8;

QY 9 LEGPEGAGTGVLEALLPILEKGVLEVTTRPFGVGLICEKIREVILPPSHMTOMDAKTEL 68